The Dissertation Committee for Jamie Lee Vernon Certifies that this is the approved version of the following dissertation:

TOWARD GROUP II INTRON-BASED GENOME TARGETING IN EUKARYOTIC CELLS

Committee:

Alan M. Lambowitz, Supervisor

Robert Krug

Paul Macdonald

Tanya Paull

Philip Tucker

TOWARD GROUP II INTRON-BASED GENOME TARGETING IN EUKARYOTIC CELLS

by

Jamie Lee Vernon, B.S.; M.S.

Dissertation

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

The University of Texas at Austin

December, 2009

Dedication

To my wonderful wife and my beautiful daughter.

To my family and friends.

To all those who gave me support when I needed it.

Acknowledgements

This dissertation would not be possible without the patience and understanding of my adviser, Dr. Alan M. Lambowitz. My family and I will forever appreciate his guidance and support during the process of writing my dissertation. When I faced the most challenging circumstances of my life, he demonstrated a tremendous sense of consideration, which gave me the opportunity to reach my goals. To him, I offer my sincerest thanks.

To my committee members who provided a great deal of support while I completed the qualifications for my Ph.D., I send my utmost thanks. Drs. Robert Krug, Paul Macdonald, Tanya Paull and Philip Tucker were instrumental in bringing my graduate experience to a successful conclusion. I will always appreciate their guidance in the scientific process. For their keen insight and perceptive recommendations, I am truly grateful.

To the members of the ICMB faculty and staff, notably Patty Lemée, Dr. William Cassady, Dr. Scott Stevens, Barbara Welch, Dr. Dean Appling and so many others, I will always remember your commitment to getting the job done and ensuring that I did mine.

To my friends from the ICMB program and the Lambowitz laboratory, especially Chad and Mireya McKee, Adam Roth, Beth Erlichman, Hari Bhaskaran, Fanglei Zhuang, Joe Hanson, Matt Nicolette, Matt Cowperthwaite and many more, I am grateful for your friendship and intellectual discourse. I look forward to a happy reunion.

Finally, to my very patient and understanding wife who stood by me when things were truly tough and to my amazing daughter who had to endure the most trying times, I love you and I thank you for just being there.

TOWARD GROUP II INTRON-BASED GENOME TARGETING IN EUKARYOTIC CELLS

Publication No._____

Jamie Lee Vernon, Ph.D.

The University of Texas at Austin, 2009

Supervisor: Alan M. Lambowitz

Mobile group II introns consist of a self-splicing RNA molecule and an intron-encoded protein with reverse transcriptase activity that function together in an RNP and catalyze the insertion of the intron into specific DNA target sites by a process known as retrohoming. The mechanism of insertion requires the intron RNA to bind and reverse splice into one strand of the DNA target site, while the intron-associated protein cleaves the opposite DNA strand and reverse transcribes the intron RNA. DNA target site recognition and binding are dependent upon base pairing between the intron RNA and the target DNA molecule. By modifying the recognition sequences in the intron RNA, group II introns can be engineered to insert into virtually any desired target DNA. Based on this technology, a novel class of commercially available group II intron-based gene targeting vectors, called targetrons, has been developed. Targetrons have been used successfully for gene targeting in a broad range of bacteria. Previously, our laboratory demonstrated that group II introns retain controllable retrohoming activity in mammalian

cells, albeit with very low targeting efficiency. However, the gene targeting capability of group II introns is not limited to direct insertion of the intron. Group II introns can also create double-strand breaks that stimulate homologous recombination. By virtue of these attributes, mobile group II introns offer great promise for applications in genetic engineering, functional genomics and gene therapy. Here I present the results of experiments in which I tested group II introns for gene targeting activities in eukaryotic cells. First, I demonstrated that group II introns injected into zebrafish (*Danio rerio*) embryos retain *in vivo* plasmid targeting activity that is enhanced by the addition of magnesium chloride and deoxynucleotides. I also verified that similar *in vivo* targeting activity is retained in *Drosophila melanogaster* embryos. Further, I describe repeated experiments in zebrafish embryos designed to target the zebrafish genome with inconclusive results. Group II introns were also delivered to cultured human cells for genome targeting. Here I present promising evidence for the ability of group II introns to stimulate homologous recombination between an exogenously introduced donor DNA molecule and the chromosome. The donor DNA was delivered either as a linearized double-stranded plasmid by electroporation or as a single stranded genome of a recombinant adeno-associated virus (AAV). In both cases, cells receiving both the group II intron RNP and the donor DNA showed more efficient integration of the donor DNA than introduction of the donor DNA alone. The studies presented here provide insight into the potential of using group II introns for future applications in gene targeting in eukaryotes.

Table of Contents

List of Tables

List of Figures

CHAPTER 1: INTRODUCTION

Group II introns are ribozymes that catalyze their own splicing reaction from precursor RNAs and are in some cases mobile elements that use their ribozyme activity to insert into DNA (Pyle & Lambowitz 2006). First discovered in the mitochondrial and chloroplast genomes of lower eukaryotes and higher plants, mobile group II introns consist of a highly structured catalytic RNA and, a multifunctional, intron-encoded protein (IEP) (Lambowitz & Zimmerly 2004). Group II introns have also been found in bacterial genomes, as well as in archa (Rest & Mindell 2003). Although only one third of organellar group II introns contain ORFs for IEPs, nearly all bacterial and archael group II introns do so (Michel & Lang 1985, Toor et al. 2001).

Sequence comparison showed that group II intron IEPs are homologous to reverse transcriptases (RTs), which are encoded by retroviruses, LTR- and non-LTR-containing retrotransposons, and telomerase (Michel & Lang 1985), and biochemical studies showed that they have RT activity (Kennell et al. 1993). Group II intron RTs typically contain four conserved domains: RT, which corresponds to the fingers and palm regions of retroviral RTs; X, which corresponds to the RT thumb; D, DNA binding; and En, DNA endonuclease (Blocker et al. 2004). The latter two domains contribute to interactions with DNA target sites during intron mobility (San Filippo and Lambowitz 2002).

The ribozyme activity of the intron RNA combined with the multifunctional IEP allows group II introns to act as mobile retroelements capable of inserting into DNA targets (Lambowitz & Zimmerly 2004). The mechanisms of group II intron retromobility and the rules for DNA target recognition have been studied extensively leading to the

development of retargetable introns that are capable of inserting into essentially any desired DNA target (Perutka et al. 2004). Sigma-Aldrich Corporation has commercialized group II intron gene targeting technology for prokaryotes under the product name of TargeTron®. Targetrons have been used in a variety of prokaryotes for gene targeting with impressive levels of success (Frazier et al. 2003, Chen et al. 2005, Yao et al. 2006, Heap et al. 2007, Shao et al. 2007, Yao & Lambowitz 2007, Rodriguez et al. 2008). The broad usage of targetrons in prokaryotes raises the question of whether group II introns can be used as a controllable gene-targeting vector for eukaryotes. Here I describe our attempts to harness the power of group II introns for uses in various metazoan systems.

1.1 Distribution and Classification of Group II Introns

Group II introns and their distribution throughout the biological world suggest a complex evolutionary history of DNA mobility, genetic transfer and evolutionary conservation, divergence and degeneration. Group II introns were originally discovered in the mitochondrial DNA (mtDNA) of lower eukaryotes, including yeast, and the chloroplast DNA (cpDNA) of alga and higher plants, where they are generally inserted within coding sequences (Lambowitz & Belfort 1993). Later, group II introns were found to be relatively common among gram-positive and gram-negative bacteria (Dai & Zimmerly 2003, Toor et al. 2001). In bacteria, group II introns are frequently inserted into intergenic sequences, mobile DNAs or plasmids, which may have promoted their dispersal throughout the prokaryotic world (Dai & Zimmerly 2002, Ichiyanagi et al. 2003). More recently, group II introns have been identified in archaeal genomes, albeit

more rarely than in bacteria (Toro 2003, Dai & Zimmerly 2002, Galagan et al. 2002, Deppenmeier et al. 2002).

The spread of group II introns is at least partly due to their ability to function as retroelements through the combined activity of the intron RNA and IEP. Extensive analysis of the phylogenetic relationships between group II IEPs identified eight clades, designated mitochondrial-like, chloroplast-like 1 and 2 and bacterial A-E (Toro et al. 2002, Zimmerly et al. 2001). Each lineage of IEP is associated with a defined RNA structural subclass. Group II intron RNAs are divided into three major subgroups (IIA, IIB, IIC) based on RNA structural variations (Michel et al. 1989, Michel & Ferat 1995, Qin & Pyle 1998, Toor et al. 2001). The most distinctive RNA structural components that define the subgroups are responsible for alignment of the splice site and the recognition of intron and exon sequences for splicing. For group IIA introns, recognition of the exons by the intron is done by the exon-binding sequences (EBS1, EBS2), which base pair with intron-binding sites (IBS1, IBS2) located in the 5' exon, and δ, which is adjacent to the EBS1 sequence and pairs with δ' corresponding to the first one or two nucleotides of the 3' exon (Michel et al. 1989). Group IIB and IIC introns have a different exon binding sequence (EBS3) in domain I that pairs with at least the first nucleotide in the 3' exon, now referred to as IBS3 (Costa et al. 2000). Groups IIA and IIB are further subdivided into subclasses: A1 and A2, B1 and B2 based on additional sequence variations (Michel et al. 1989, Michel & Ferat 1995, Qin & Pyle 1998).

Despite the widespread presence of group II introns in organelles and prokaryotes, only a few group II introns have been identified in the mitochondrial genes of animals,

each from early diverging metazoan lineages including the annelid worm, *Nephtys*, and the placozoan, *Trichoplax adhaerens* (Valles et al. 2008, Dellaporta et al. 2006). Group II introns have yet to be identified in any metazoan animal organellar or nuclear genome. However, due to significant similarities in splicing mechanisms and splicing products, group II introns are believed to be the evolutionary predecessors of the spliceosomal introns of animals and other eukaryotes (Cech 1986). An indication that group II introns might have given rise to spliceosomal introns is the finding that some group II introns in plant mitochondria and chloroplasts are transcribed in two or three segments that re-associate via tertiary interactions and carry out "trans-splicing" (Bonen 1993). The ability of such independently transcribed segments to functionally reassociate supports the hypothesis that group II intron domains may have evolved into the snRNAs that reassociate to form the spliceosome (Sharp 1985).

One explanation for the evolution of group II introns into spliceosomal introns is that shortly after the origin of the mitochondrion via endosymbiotic absorption of an α-proteo-bacterium and the chloroplast from cyanobacteria, group II introns escaped these organelles and invaded the host cell chromosomes, subsequently spreading throughout the genome (Palmer & Logsdon 1991, Cavalier-Smith 1991, Martin & Koonin 2006). Martin and Koonin (2006) expanded upon this hypothesis by suggesting that the spread of group II introns and their mutational decay into spliceosomal introns created a strong selective pressure for independent compartmentalization of the splicing and translation pathways giving rise to the need for the nucleus-cytosol separation. The absence of functional group II introns in eukaryotic genomes suggests that eukaryotes have avoided

recent group II intron invasions since the emergence of spliceosomal introns, but does not preclude the possibility that group II introns could be artificially introduced into eukaryotes for technological purposes.

1.2 RNA Structure and Self-splicing of Group II Introns

Group II intron RNAs catalyze splicing by folding into highly conserved secondary and tertiary structures that form an active site, just as in a protein enzyme (Pyle & Lambowitz 2006). Despite having little primary sequence conservation, all group II introns exhibit a conserved secondary structure that consists of a central wheel with six double helical domains (DI-DVI) radiating outward (Figure 1.1). These six domains fold to form a tertiary structure that catalyzes two magnesium-dependent transesterification reactions; 1) the bulged A residue in DIV creates the lariat structure by reacting with the first nucleotide of the intron to form a 2'-5' phosphodiester linkage and releasing the 5' exon, 2) the 3' hydroxyl of the free 5' exon attacks the phosphodiester bond at the 3' splice site, thereby ligating the exons (Figure 1.2). The products of the splicing reactions are ligated exons and an excised intron lariat (Peebles et al. 1986, Schmelzer & Schweyen 1986, van der Veen et al. 1986).

An alternative-splicing pathway has also been observed in which a water molecule catalyzes the hydrolysis of the 5' splice site. The product of this reaction, a linear instead of a lariat intron, was shown to occur in self-splicing reactions under non-physiological conditions (Zhuang et al. 2009, Roitzsch & Pyle 2009, Pyle et al. 2006) (Figure 1.2). Branch point mutants, missing the bulged A residue from DVI, of the yeast mitochondrial (mt) group II intron aI5γ can splice *in vivo* to form the linear intron (Podar et al. 1998).

While other introns have been observed to undergo splicing *in vivo* by mechanisms that yield linear intron RNA, it is currently thought the most common mechanism for group II intron splicing is the lariat-forming pathway (Podar et al. 1998, Vogel and Börner 2002).

The minimal catalytic core for the group II intron splicing reaction consists of DI and DV (Michel et al. 1989). DI contains exon-binding sequences (EBS1 and EBS2) that form classical Watson-Crick base pairs with specific nucleotides in the exons (intron binding sequences, IBS1 and IBS2) (Figure 1.3) (Jacquier & Michel 1987, Michel & Jacquier 1987). Base pairing interactions between the intron and 5' exon are essential for RNA splicing, but differ somewhat for the different intron classes. DI also contains sequences (δ or EBS3) that contribute to splicing by base pairing to the flanking 3' exon sequence (Michel & Ferat 1995; Costa et al. 2000). DV, the most conserved hairpin sequence among group II introns, docks with DI to form essential active-site structures (Qin & Pyle 1998).

While they do not participate in catalysis of the splicing reaction, DII and DIII contribute to RNA folding and catalytic efficiency. Splicing reactions can proceed under high salt conditions when DII and DIII are removed, but the efficiency of the second step of the splicing reaction is significantly reduced (Fedorova et al. 2003, Michel & Ferat 1995). DIV is the least conserved domain of group II introns and is largely dispensable for ribozyme activity. However, DIV is significant because it contains the ORF sequence for the IEP, which is notably involved in the splicing reaction.

1.3 Group II Intron-Encoded Proteins and Protein-Assisted Splicing

Group II intron IEPs are required for both RNA splicing and intron mobility (Michel & Ferat 1995). Whereas some group II introns are able to self-splice *in vitro*, this self-splicing generally requires non-physiological conditions, such as high salt, Mg^{2+}, and/or temperature, and does not occur efficiently. The ribozyme and the IEP have coevolved to such a degree that the protein is required *in vivo* to help the intron RNA fold into a catalytically active structure (Michel & Ferat, 1995, Lambowitz & Zimmerly 2004, Lambowitz & Belfort, 1993, Zimmerly et al. 1999). More importantly for dispersion of group II introns, following splicing the protein remains associated with the excised intron RNA as a ribonucleoprotein, RNP, and participates in intron mobility reactions (Lambowitz & Zimmerly 2004).

Group II intron IEPs have sequence homology to the RTs of retroviruses and retrotransposable elements (Michel & Lang, 1985). One hypothesis for the presence of RTs in group II introns posits the insertion of a pre-existing RT ORF, such as a retron RT, into a self-splicing RNA to form a new type of retrotransposable element (Lambowitz & Belfort, 1993, Wank et al. 1999). Since the similarity to reverse transcriptases was identified, additional domains have been discovered in the coding sequences of group II intron IEPs, including X, thumb/maturase; D, DNA binding, and En, DNA endonuclease (Figure 1.4). Each of these domains contributes to the activities of the protein.

The RT and X/thumb domains, which share conserved sequences with HIV-1 RT, bind specifically to the intron RNA and stabilize the catalytically active RNA structure

that is required for RNA splicing (Figure 1.4) (Saldanha et al. 1999, Wank et al. 1999, Matsuura et al. 2001, Noah & Lambowitz 2003, Blocker et al. 2005). RNA splicing does not require either the D or En domains. In fact, more than half of bacterial IEPs lack an endouclease domain (En) and many are also missing a readily identifiable DNA-binding domain (D) (Martinez-Abarca et al. 2000, Zimmerly et al. 2001, Belfort 2003, Toro 2003, Lambowitz & Zimmerly 2004). Analysis of the En and D domains of LtrA, revealed that the two domains may have been acquired together from an endonuclease VII-type protein (San Filippo & Lambowitz 2002). Certain lineages of group II introns appear to have lost the D and En domains because they were deleterious to the host (San Filippo & Lambowitz 2002). When present, however, the En and D domains, in conjunction with the RT domain, provide critical enzymatic activities that contribute to the mobility of group II introns.

1.4 Mechanism of Group II Intron Mobility

The distribution of group II introns throughout the biological world is dependent upon the ability of the introns to mobilize in permissive cellular environments. The principal means by which group II introns spread are called retrohoming and retrotransposition. Retrohoming is the process by which the intron RNA inserts into a specific DNA target sequence, often the ligated exon junction of an intronless allele, referred to as the homing site. Retrotransposition occurs at a lower frequency and involves insertion into a non-cognate DNA target site that resembles the normal homing site. The latter pathway serves as a major contributor to distribution of group II introns into novel DNA sites through ectopic integrations (Toro et al. 2007, Pyle & Lambowitz

2006, Lambowitz & Zimmerly 2004). Both retrohoming and retrotransposition use essentially the same mechanism for integration of the intron into DNA sites.

The mechanism for integration of the group II intron into a DNA target site is referred to as target DNA-primed reverse transcription (TPRT) and is analogous in some respects to the TPRT mechanism used by non-LTR retrotransposons (Zimmerly et al. 1995, Eickbush 1999). The process of target site integration is dependent upon the RNA and the IEP, which remain bound together following the forward splicing reaction to form a ribonucleoprotein complex (RNP). In order to initiate the mobility pathway, the RNP must first identify a DNA target site. Group II intron RNPs recognize relatively long (30-35) stretches of nucleotides via DNA-binding regions of the IEP and base-pairing interactions between the intron RNA and the DNA target site (Guo et al 1997, Guo et al. 2000, Mohr et al. 2000, Singh & Lambowitz 2001). For the Ll.LtrB intron, the RNA base pairs with the DNA target site from position -12 to +2. The IEP recognizes a small number of nucleotide residues in the regions flanking the RNA-DNA base pairs. Following recognition of the target site, the intron RNA reverse splices into the top strand of the DNA (Lambowitz & Zimmerly 2004, Lambowitz & Pyle 2006). Like some non-LTR retrotransposons, group II intron RNPs that have an intact En domain can generate a specific cleavage at their target sites (Malik & Eickbush 1999, San Filippo & Lambowitz 2002, Lambowitz & Zimmerly 2004). Nevertheless, the Ll.ltrB intron is able to utilize both En-dependent and En-independent means of mobilization (Zhong & Lambowitz 2003, Coros et al 2005). When employing the En-dependent mechanism, the En domain of the IEP, which contains an H-N-H DNA endonuclease motif, cleaves the bottom

strand of the DNA target site between nucleotides +9 and +10 (San Filippo & Lambowitz 2002). Cleavage of the bottom strand creates a free 3' hydroxyl group that functions as the primer for cDNA synthesis of the intron RNA by the RT domain of the IEP (Eskes et al. 1997, Cousineau et al. 1998, Eskes et al. 2000). In the En-independent mechanism, the target site recognition and reverse splicing reaction are essentially the same but a nascent strand at a DNA replication fork is used to prime reverse transcription of the inserted intron RNA (Zhong & Lambowitz 2003). In both cases, once the full-length cDNA has been generated, subsequent repair of the cDNA junction and removal and replacement of the intron RNA are performed by host enzymes (Cousineau et al. 1998, Smith et al. 2005, Eskes et al. 1997, Eskes et al. 2000).

While the above mentioned pathways are specific to lariat group II introns, a novel mobility pathway for linear group II introns has recently been proposed (Zhuang et al. 2009, Roitzsch & Pyle 2009). Here, it was shown that linear group II introns can carry out the first step of the mobility process by reverse splicing into DNA targets followed by reverse transcription of the intron RNA (Mastroianni et al. 2008). The resulting cDNA was integrated into the recipient DNA by host DNA repair pathways (Mastroianni et al. 2008, Zhuang et al. 2009). These three mobility pathways offer a broad range of potential applications for group II introns in gene targeting.

1.5 Use of Group II Introns for Gene Targeting and Genetic Engineering in Prokaryotes

Since intron mobility was originally demonstrated using the yeast and *K. lactis* *cox*I-I1 introns, a great deal of work has been done to understand the processes involved (Meunier et al. 1990, Skelly et al. 1991). Once the mechanisms for recognizing the DNA

target site and integration were clearly defined, a new era of group II intron engineering emerged. The earliest attempts at engineering group II introns were designed to alter their targeting specificity by modifying the RNA and DNA base-pairing interactions of the yeast mitochondrial aI1 and aI2 and the *Lactococcus lactis*, Ll.LtrB intron RNAs (Guo et al. 1997, Eskes et al. 1997, Matsuura et al. 1997). The greatest success was achieved using Ll.LtrB intron due to the development of an efficient *E. coli* expression system for this intron (Matsuura et al. 1997), and it has become the best characterized and most widely used group II intron for genetic engineering experiments.

Shortly after the first experiments involving modification of base-pairing interactions were performed, detailed DNA target-site recognition rules were established for the Ll.LtrB intron (Mohr et al. 2000, Guo et al. 2000, Zhong et al. 2003). It was determined that group II introns could be retargeted to insert efficiently into desired DNA targets simply by modifying the intron RNA (Guo et al. 2000). A randomized intron library containing a retrotransposition-activated genetic marker (RAM) was used to generate a database of introns and target sites along with their integration efficiencies (Zhong et al. 2003). The information gathered in these studies was used to construct an algorithm that designs introns to target a given gene with high efficiency and a new prokaryotic gene targeting technology was born (Perutka et al. 2004). This technology allowed for development of programmable bacterial gene targeting vectors, called "targetrons." Targetron technology has been used to successfully disrupt and deliver genes in a variety of medically and commercially relevant strains of gram-positive and gram-negative bacteria (Frazier et al. 2003, Shao et al. 2007). By introducing retargeted

Ll.LtrB introns into a strain of bacteria via a donor plasmid expressing the intron and IEP, both conditional and non-conditional gene disruptions can be generated, depending on the orientation of the intron insertion and expression of the IEP (Guo et al. 2000, Karberg et al. 2001, Frazier et al. 2003, Yao et al. 2006). Clearly, group II introns provide an impressive set of tools for prokaryotic genetic engineering. Similar success in eukaryotes would make obsolete many of the current eukaryotic genetic engineering tools.

1.6 Current Methods of Gene Targeting and Genetic Engineering in Eukaryotes

In many cases, current eukaryotic gene-targeting methods are organism-specific and are often limited by inherent disadvantages. For the mammalian model organism, the mouse, a commonly used gene silencing process called RNA interference relies on short dsRNA molecules called siRNAs (Paddison et al. 2002). A cytoplasmic ribonuclease called dicer processes the siRNAs from longer precursors. The antisense strand of the siRNA remains associated as the effector for the RNA-induced silencing complex (RISC) to recognize and cleave a complementary messenger RNA, targeting it for degradation (Meister et al. 2004). Application of siRNAs has been effective, in many respects, for determining the role of specific genes in the mouse genome, but interpretation of siRNA studies are complicated by the fact that this method creates knock-downs, not knock-outs (Mittal 2004). Also, the siRNA oligos used in these experiments are unstable and must be constantly replenished for long-lasting gene silencing (Sandy et al. 2005). High intracellular concentrations of siRNAs can be maintained by expressing them from stably transfected plasmids or from integrated viral vectors. These types of expression

constructs obviously run the risk of integrating uncontrollably into the genome with unintended consequences. In recent years, the need to develop efficient ways of designing, identifying and delivering effective siRNAs has also become more important. Until these complications are addressed, the full impact of siRNA technology will remain at large.

While siRNAs are widely used in the mouse, the most common method of gene targeting is the site-specific integration of exogenous DNA by homologous recombination (HR). The advantage of HR technology over siRNAs is that HR creates germ-line mutations that are maintainable in cell culture and transferable to animals. Generally, this involves the introduction into cultured mouse stem cells of foreign double-stranded DNA that is engineered to contain regions of sequence homology to the chromosome (Wang & Zhou 2003). The donor DNA molecule used for HR-based gene targeting traditionally includes a selectable marker inserted in such a way as to disrupt the target gene upon integration (Figure 1.5). Delivery of the donor DNA to the nucleus by microinjection, electroporation or lipid-based transfection triggers the homologous recombination machinery to align the regions of homology to the chromosome and incorporate the exogenous DNA into the target site. Cells containing targeting events are isolated using selectable markers expressed from the integrated donor DNA. To further aid in the selection process, donor DNAs can be designed to express negative selection markers from non-integrated or randomly integrated copies. Because HR is a highly inefficient process (10^{-6} integrations/cell), the screening methods can be labor-intensive

and cost-prohibitive. Further, the need to create and introduce transgenic material into the cell always leads to concerns about off-target effects.

More recently the use of adeno-associated virus for delivery of the donor DNA in HR-based gene targeting has become an attractive alternative to double-stranded plasmid DNA. The genome of AAV, a human parvovirus with broad cellular tropism, is encapsidated as a single-stranded DNA. Recombinant AAV (rAAV) can be engineered by exchanging the viral genes located between the inverted terminal repeat sequences (ITRs) with up to 4.7 kb of target-derived replacement DNA (Porteus et al. 2003, Russell & Hirata 1998, Samulski et al. 1989). rAAV gene targeting vectors provide gene disruptions 3-4 orders of magnitude more efficiently than those achieved by conventional targeting approaches using double-stranded DNA (Yan et al. 2009). Although the mechanism of rAAV homologous recombination is poorly understood, it is assumed that it occurs through processes similar to conventional gene targeting. The advantages that rAAV has over other targeting strategies may reside in the single stranded genome of the virus, which is flanked by the highly ordered secondary structure of the ITRs. These structures may be highly efficient at binding DNA repair enzymes, thereby more easily integrating into the genome (Yan et al. 2009). While the highly efficient gene targeting of rAAV is attractive, the need for extremely high MOIs to achieve these results limits the use of rAAV to non-therapeutic purposes (Porteus et al. 2003). Further, the potential for random integration is currently too high at even low MOIs. Clearly, there is a need for an improved gene targeting strategy for mice. Similar problems exist for gene targeting strategies in other model organisms, like zebrafish.

Unlike in mouse gene targeting experiments, which are most often carried out in cell culture, injecting the gene targeting effectors directly into early stage embryos is the usual protocol for zebrafish gene targeting. Techniques that are commonly used in mouse, such as homologous recombination and RNAi, are not routinely available for zebrafish (Deiters & Yoder 2006). First, homologous recombination is not the preferred mechanism for integration of exogenous DNA during embryonic development in zebrafish. Although the cellular machinery for homologous repair is present in zebrafish embryos, double-strand breaks tend to be predominantly repaired by DNA end-joining mechanisms, so exogenously introduced linear DNAs tend to integrate randomly into the genome (Hagmann et al. 1998). Secondly, the effective concentration necessary for traditional RNAi-based targeting, in mouse, is toxic to zebrafish embryos. A commonly used and reasonably successful technique for gene disruption in zebrafish relies on the mutagenic capacity of retroviruses (Amsterdam et al. 1999). Like the advances made in *Drosophila melanogaster* using P elements (Cooley et al. 1988), insertional mutagenesis using pseudo-typed retroviruses in zebrafish speeds the process of cloning mutant genes significantly, compared to chemical mutagenesis (Amsterdam et al. 1999). While some labs have been very successful using insertional mutagenesis in zebrafish, the process requires large-scale screening, is not target-specific, and occasionally results in secondary mutations that induce additional phenotypes (Iversen & Newbry 2005).

The most commonly used method of studying gene function in zebrafish utilizes modified gene-targeted RNA oligonucleotides, called morpholinos (Nasevicius & Ekker 2000). The term morpholino refers to phosphoramidite morpholino oligomers that have

an advantage over traditional antisense RNA oligos because they are inert and do not elicit toxic effects in zebrafish cells. However, as is true for siRNAs in mice, the morpholino oligos do not result in complete knockouts of gene function, but rather lead to a knockdown of gene expression. Further, a common criticism of morpholino gene targeting is that it is not 100% effective (Iversen & Newbry 2005). In fact, morpholino gene targeting is unsuccessful, as often as, 26% of the time (Iversen & Newbry 2005). The non-specific effects associated with morpholinos have not been fully characterized and there is little impact on genes expressed in embryos older than two days (Eisen & Smith 2008, Iversen & Newbry 2005).

A promising method of gene targeting, known as tilling (Targeting Induced Local Lesions in Genomes), recently emerged as the next promising tool for studying zebrafish genetics. Tilling is a hybrid of reverse and classical genetic approaches whereby genomic DNA from a large library of ENU-mutagenized zebrafish is screened for rare mutations in genes of interest. It is the screening method that makes this technology more effective than traditional genetics because mutations in specific genes can be detected directly in the genomic DNA of heterozygous or homozygous individuals, irrespective of any phenotypes they may cause (Moens et al. 2008). Unfortunately, because this is a PCR-based screening method that relies on rare chemical-induced mutations, amplicons that contain a high frequency of naturally occurring single nucleotide polymorphisms (SNPs) can complicate the screening process (Barkley & Wang 2008). The TILLING method also requires a great deal of experience to develop the ability to distinguish false positives and false negatives from actual mutants. Further,

it can be labor intensive when not using robotic equipment to process the large numbers of samples involved and, thusly, the recommended equipment for TILLING is often cost prohibitive for smaller labs.

A gene targeting technique that offers promise for uses in several different model organisms capitalizes on the advances made in the field of chimeric restriction enzymes, specifically zinc-finger nucleases (ZFN) (Porteus & Carroll 2005). Here, a non-specific restriction endonuclease (*Fok*I) domain is fused to a DNA binding motif, in this case a zinc-finger DNA binding domain (Chandrasegaran & Smith 1999). Since the recognition specificity of zinc fingers can be manipulated experimentally, the chimeric nuclease can be engineered to recognize specific sequences within the genome. Using these re-targetable endonucleases, researchers can take advantage of various cellular repair mechanisms to carry out genome engineering.

By introducing a double-strand break at a specific target site, one can expect multiple pathways to be involved in repairing, editing or mutating the genome. For example, repair by the inherently mutagenic non-homologous end-joining (NHEJ) pathway can disrupt a gene that has suffered a double-strand break, either naturally or artificially (van Gent et al. 2001). Additionally, the introduction of a double-strand break at a specific site in a gene can stimulate the homologous recombination pathway, which naturally repairs DSBs by using the undamaged sister chromatid as a template (van Gent et al. 2001). Alternatively, the introduction of an exogenous DNA containing sufficient homology to the region of the double-strand break can co-opt the host cell's DSB repair pathway to incorporate the foreign DNA at the cleavage site. In fact, ZFNs have been

shown to stimulate gene targeting by directing the homologous recombination machinery to site-specifically integrate exogenous DNAs (Cathomen & Joung 2008, Porteus & Carroll 2005). The combination of artificially introduced double-strand breaks by ZFNs with HR-based incorporation of donor DNA molecules increases targeting efficiencies from 1 in 10^5 cells to as much as 39% in certain cell lines (Cathomen & Joung 2008). Thusly, the use of ZFNs for gene targeting has been extended to a variety of model organisms including mouse, *Xenopus*, zebrafish and *Drosophila*.

As powerful as the zinc-finger methodology appears to be, it has the disadvantage of requiring protein engineering in order to retarget the nuclease to new target sites. This is a notable drawback because it limits the rate and extent to which these enzymes can be modified. Further, each of the two existing methods of engineering ZFN has inherent problems. Modular assembly, the joining together of single zinc fingers with pre-characterized specificities, has an efficacy rate for making functional ZFN pairs that is less than 6% and can yield ZFNs with low activities and high toxicities (Ramirez et al. 2008, Bae et al. 2003, Beerli & Barbas 2002, Liu et al. 2002, Mandell & Barbas 2006, Segal et al. 2003). Combinatorial selection-based methods that yield multifinger domains possessing high activities and low toxicities in human cells require construction and interrogation of large randomized libraries that make them intractable in most labs that lack the necessary expertise (Cornu et al. 2008, Pruett-Miller et al. 2008). No matter the promise or widespread use of existing gene targeting technologies, it seems the disadvantages for each method leave room for new methods that address the existing limitations.

1.7 Promise and Significance of Eukaryotic Gene Targeting Using Group II Introns

Clearly, a breakthrough in group II intron-based gene targeting technology for eukaryotes would be highly desirable. Because we are proposing that group II introns possess many attractive characteristics regarding their potential applications in gene targeting and transgenesis, it seems necessary to compare their use to the most promising new gene targeting technologies currently being used today. First, the ease of retargeting group II introns greatly exceeds the complicated methods for retargeting ZFNs and offers much more potential versatility over other gene targeting methods that rely on random events. The mechanism of group II intron-based gene targeting also offers a wide range of biochemical activities that can be exploited for gene targeting. First, the most obvious mode of gene disruption by group II introns is direct integration of the intron at the target site. This provides the opportunity to create conditional and non-conditional gene knock-outs. As previously mentioned, integration of the intron in the sense strand of the coding sequence for the target gene allows for the addition of the IEP to facilitate removal of the intron from the transcript. This results in an intact wild-type transcript that can undergo normal processing and translation. Secondly, group II introns can create site-specific double-strand breaks (Karberg et al. 2001, Mastoianni et al. 2008). By cutting the genome, group II introns can cause gene disruptions by either NHEJ or homologous recombination-dependent mechanisms. Finally, group II introns offer the extraordinary ability to deliver cargo genes to a chosen target site via insertion of genetic information into DIV of the intron (Karberg et al. 2001, Frazier et al. 2003). A vector, like the group II intron, that could deliver a transgene to a specific site in the genome would be an

invaluable tool. Not only could this facilitate transgenic expression of a gene of interest from a chosen location, a gene could also simultaneously be disrupted. Significant success using group II introns in eukaryotes could pave the way for an entirely new approach to eukaryotic genetic engineering.

1.8 Overview of Dissertation Research

This dissertation research has the major theme of advancing the use of group II introns toward gene targeting in eukaryotic cells. I addressed two major subtopics: 1) examination of group II intron activity in eukaryotic embryos including zebrafish (*Danio rerio*) and *Drosophila melanogaster* (plasmid targeting and genome targeting in zebrafish, plasmid targeting in *Drosophila*) and 2) gene targeting in mammalian cells utilizing group II intron RNPs as site-specific endonucleases to stimulate homologous recombination of exogenously introduced donor DNA molecules (either a linearized double-stranded plasmid DNA or a recombinant single stranded adeno-asssociated virus, AAV, genome). The results of this work will be summarized in two chapters followed by a chapter describing the Materials and Methods.

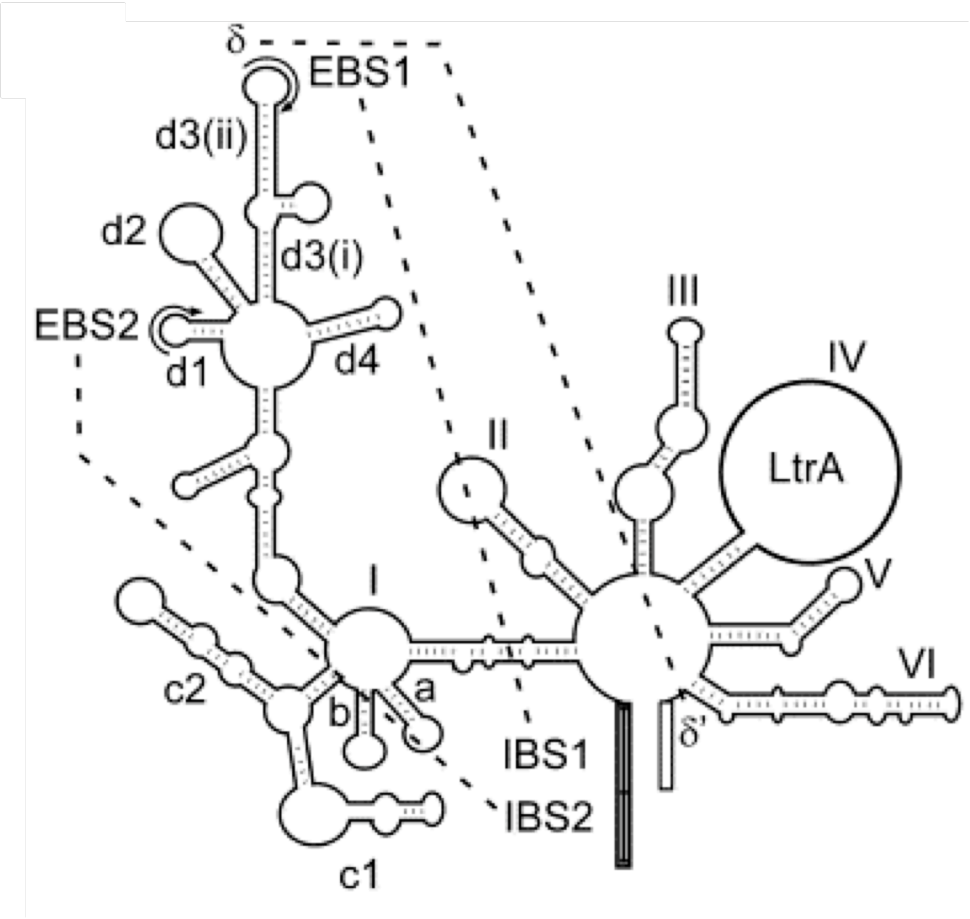


Figure 1.1 Conserved Secondary Structure and Predicted Tertiary Interactions of the Model Group IIA Intron, Ll.LtrB.

The conserved secondary structure common to group II introns is composed of six double-helical domains (I-VI) radiating from a central wheel. In this diagram, the structure of domain IV, which contains an ORF in some introns, is simplified to emphasize the overall structure of the intron. The EBS1, EBS2 and δ sequences are labeled in domain I and dotted lines reflect the base pairing interactions with the IBS1, IBS2 and δ' sequences in the 5' and 3' exons that are required for intron splicing. A bulged A residue in DIV creates the lariat structure during the splicing reaction by reacting with the first nucleotide of the intron to form a 2'-5' phosphodiester linkage. Adapted from Perutka et al., 2004.

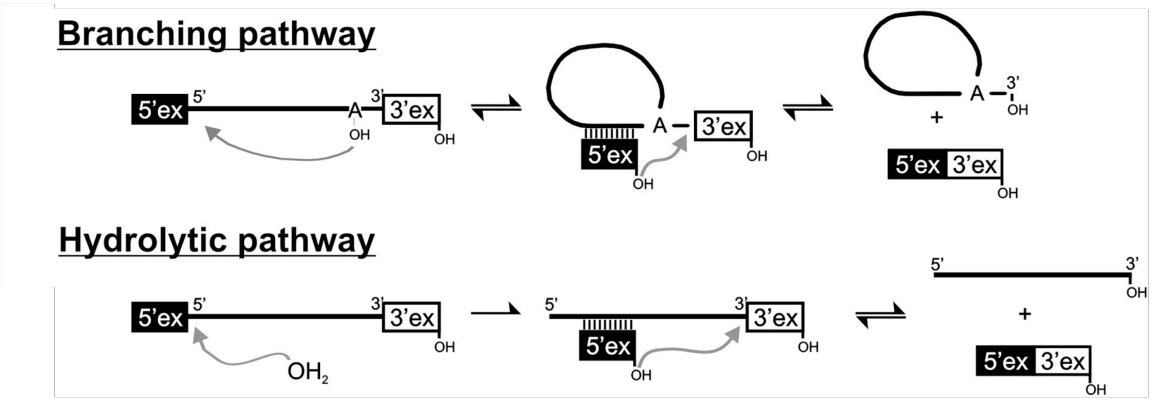


Figure 1.2 Splicing Mechanisms for Group II Introns.

The self-splicing reaction of group II introns can follow two different pathways. The two-step process can produce both linear and lariat introns. The lariat pathway begins when the 2′-OH group of the bulged adenosine in DVI reacts with the first nucleotide of the intron to form a 2'-5' phosphodiester bond releasing the 5' exon. The 3' hydroxyl then attacks the phosphodiester bond at the 3' splice site to create the lariat form of the intron. The alternative splicing pathway requires a water molecule to catalyze the hydrolysis of the 5' splice site without forming a branched structure. The 3' exon is removed from the intron in a manner similar to the lariat-forming pathway when the 5' exon reacts with it to ligate the exons. Adapted from Roitzsch & Pyle, 2009.

(a) (b)

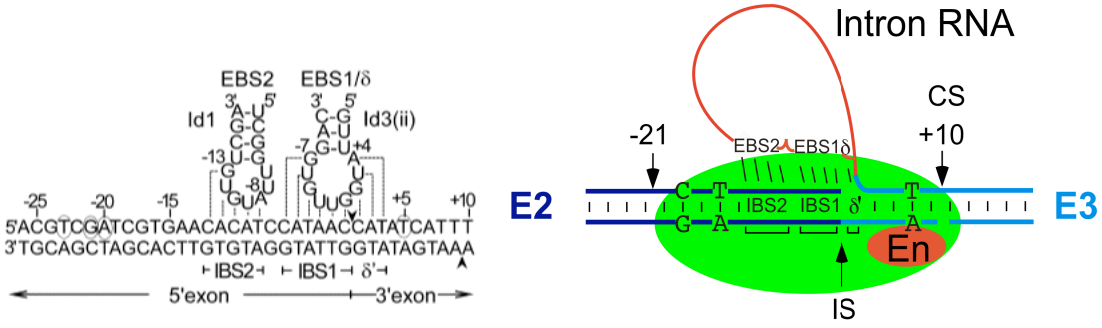


Figure 1.3 Base-Pairing Interactions Between the Intron and the Exons.

(a) The exon-binding sequences (EBS1 and EBS2) in DI form classical Watson-Crick base pairs with specific nucleotides in the exons (intron-binding sequences, IBS1 and IBS2). Base-pairing interactions are essential for RNA splicing. DI also contains sequences (δ and EBS3) that contribute to splicing by base pairing to the flanking 3' exon sequence (Michel & Ferat, 1995, Costa et al., 2000). Adapted from Perutka et al., 2004.

(b) Base pairing interactions (EBS1/IBS1, EBS2/IBS2 and d/d'), lariat structure, insertion site (IS) for the intron RNA and the cleavage site (CS) at nucleotide +10 for the endonuclease domain of the intron-encoded protein (En).

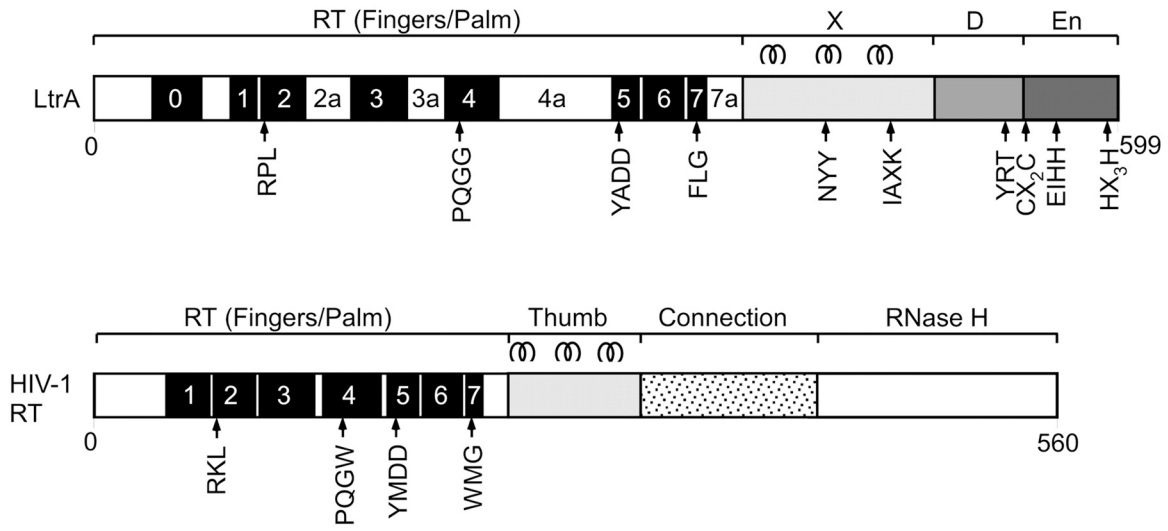


Figure 1.4 Domains of the Group II Intron IEP and Comparison to HIV-1 RT.

The group II intron IEP contains several important domains including the RT region, which shares conserved sequence motifs RT1–RT7, found in the fingers and palm, with the HIV-1 RT. The N-terminal RT domain of HIV-1 RT precedes the thumb, connection, and RNase H domains, while the RT domain of LtrA is located upstream of the thumb (X), DNA-binding (D), and DNA endonuclease (En) domains. Adapted from Blocker *et al*. 2005.

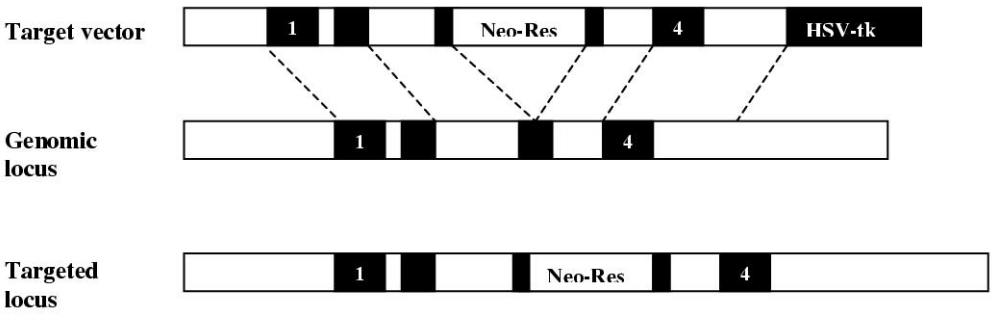


Figure 1.5 Traditional Targeting Vector for Mouse ES Cells.

Targeting vectors most commonly used for homologous recombination-based gene targeting in mouse ES cells include homology arms that flank a selectable marker. The marker disrupts the target gene upon insertion. The targeting vector can be designed to express negative selection markers, as seen here (HSV-tk), to aid in the selection process by counter selecting random integrations. Adapted from Wang & Zhou, 2003.

CHAPTER 2: GROUP II INTRON-BASED GENE TARGETING IN EUKARYOTIC EMBRYOS

In the field of eukaryotic molecular genetics, legions of scientists have sought to apply reverse genetics approaches to solve problems unanswered by classical genetic methods. The strength of reverse genetics is the ability to disrupt a specific gene of interest without the need to screen through a vast number of random mutations. Another method of studying genes by reverse genetics, known as transgenesis, allows researchers to learn about the function of a specific gene of interest by controlling the spatial and temporal expression of the gene. By using transgenesis to control the expression of a particular gene or using gene targeting to disrupt the expression of a given gene, scientists have been able to gain knowledge that was previously indiscernible using basic genetic tools. The combined benefits of gene targeting and transgenesis have influenced nearly every field of biological research from evolution and development to cancer biology. Although a broad range of techniques have been developed for transgenesis and gene targeting in eukaryotes, each has distinct disadvantages that invite the development of improved technologies. Today, one of the greatest challenges in molecular genetics is the development of an efficient, reliable, controllable, cross-species gene targeting method.

To this end, efforts have been made to prove the practicality of group II introns for eukaryotic gene targeting. Initially, it had to be determined whether group II introns could recognize and target eukaryotic sequences (Guo et al. 2001). Group II introns were designed to insert into sequences from the HIV-1 provirus and the human gene encoding

the HIV-1 co-receptor CCR5. Using the *E. coli* system developed to test the retargetability of group II introns, it was proven that group II introns could efficiently integrate into both the HIV-1 provirus and CCR5 DNA target sites in *E. coli* plasmid assays (Guo et al. 2001).

Subsequently, the *E. coli* assay was also used to show that a mutant β-globin gene could be repaired by group II intron-based delivery of functional exons into a plasmid vector containing the sequence of the deactivated β-globin gene (Jones et al. 2005). Functional exons 2 and 3 of the β-globin gene were cloned into DIV of a group II intron that had been retargeted to insert into the first intron of the β-globin gene. To ensure proper repair of the gene, a canonical 3' splice acceptor site was cloned upstream of the exons and a polyA signal was cloned immediately downstream of exon 3 (Jones et al. 2005). Following successful intron integration into the β-globin sequence using the *E. coli* system, the plasmid DNA carrying the corrected version of the β-globin gene was transfected into mammalian cells. The cells expressed the RNA and protein corresponding to wild-type β-globin.

Whereas the previously mentioned experiments were performed in *E. coli*, Guo et al. 2001 demonstrated for the first time that group II introns retained activity in human cells and could successfully integrate into plasmid-borne therapeutically relevant DNA target sites (Guo et al. 2001). In these experiments, *in vitro*-prepared RNPs designed to target the HIV-1 co-receptor CCR5 were introduced separately from the plasmid DNA targets using liposome-mediated transfection (Guo et al. 2001). The RNPs were able to locate and insert into the plasmids with limited efficiency, detectable only by nested

PCR. Overcoming the low targeting efficiencies of those early experiments is essential for developing group II introns into a viable tool for eukaryotic gene targeting. We felt that correctable technical issues, such as inefficient delivery or disruption of the RNP complex by the transfection agent, were the likely reasons for the low targeting efficiencies (Guo et al. 2000). Regardless, the need to improve the targeting efficiencies was obvious, and the potential impact made the investment in developing the technology worthy of the time and effort.

To address the issue of targeting efficiency, we began by considering the possible problem areas affecting group II intron activity in eukaryotic cells. In addition to addressing the technical issues related to RNP delivery, it quickly became evident that the *in vivo* magnesium concentration was suboptimal (Romani & Maguire 2002). Based on *in vitro* targeting experiments, the optimal magnesium concentration was determined to be approximately 10 mM (Saldanha et al. 1999). The high level of targeting in prokaryotes even at suboptimal magnesium levels in bacteria suggests that host factors may play a role in the targeting pathway. However, to date, few examples of helpful prokaryotic host factors have been confirmed to have a significant effect on gene targeting. Accordingly, we set out to determine conditions under which plasmid targeting could be improved in eukaryotes. We hypothesized that addition of magnesium to the cellular environment would increase targeting efficiencies. We also considered the need for efficient localization of group II intron RNPs to the nucleus and improved penetration of the eukaryotic chromatin structure, which could limit access to genomic targets. Finally, we reflected on the possibility that the cellular environment could be hostile such

that introduction of RNPs into the nucleus or cytoplasm would elicit a response that would neutralize the RNP activity. In order to address these issues, we designed experiments to directly manage many of these potential obstacles.

The first experiments that I performed included the injection of group II intron RNPs into live zebrafish embryos. Injecting *in vitro*-prepared RNPs allowed us to circumvent obstacles to expression and assembly of the RNP complex in the cell. By eliminating the need to express the intron *in vivo*, we could bypass transcriptional and translational regulatory pathways that might inhibit RNP formation. Injections also allowed us to control the amounts of RNPs introduced into the cell and regulate the intracellular magnesium levels. The latter allowed for systematic determination of the optimal *in vivo* magnesium concentrations for efficient gene targeting in zebrafish embryos. Finally, injecting into actively dividing embryos exposed the RNP to a dynamic cellular environment that presented opportunities for access to the chromosomal DNA, such as nuclear envelope breakdown and periods of DNA replication .

2.1 Reconstitution of Group II Intron RNPs for Microinjection and TPRT Assays

In order to ascertain the targeting capabilities of group II introns in eukaryotic cells, it was necessary to make RNPs that were purified and concentrated enough for microinjections. The development of an efficient *Eschericia coli* expression system for the *L. lactis* group II intron Ll.LtrB allowed us to express and purify the components of the Ll.LtrB group II intron RNP (Matsuura et al. 1997; Saldanha et al. 1999). Matsuura et al. (1997) demonstrated that both the intron and the IEP could be expressed *in vivo* at sufficient levels for purification of either the LtrA protein alone or the fully reconstituted

RNP. Protein expressed in *E. coli* can also be purified and combined with *in vitro* transcribed intron RNA following the self-splicing reaction to form active RNP complexes. Testing of purified RNPs, prepared in this manner, proved that LtrA has RT, maturase and site-specific endonuclease activity (Matsuura et al. 1997). Saldanha et al. (1999) showed that *in vitro*-reconstituted RNPs, composed of purified LtrA and the excised lariat intron RNA, are able to carry out *in vitro* target-primed reverse transcription (TPRT) reactions, under defined reaction conditions. The ability to perform *in vitro* TPRT demonstrated that purified reconstituted RNPs are capable of locating the target site, reverse splicing and carrying out reverse transcription with a high degree of specificity and efficiency.

For the experiments described here, I applied methods similar to those established for purification of LtrA and *in vitro* reconstitution of the RNPs (Matsuura et al. 1997, Saldanha et al. 1999). For expression of the LtrA protein, *E. coli* cells were transformed with the plasmid, pIMP-1P. The plasmid contains the LtrA ORF cloned downstream of a *tac* promoter and ϕ10 Shine Dalgarno sequence in the pCYB2 vector backbone (New England Biolabs). LtrA was expressed as a fusion protein with a C-terminal tag containing an intein-linked chitin-binding domain, enabling LtrA purification via a chitin-affinity column, followed by intein-cleavage. For the *in vitro* transcription of wild-type Ll.LtrB-ΔORF intron RNA, I used the pACD2 intron-donor plasmid, which contains the Ll.LtrB-ΔORF intron with an additional T7 promoter in DIV. The intron contained the wild-type (WT) EBS sequences so it would recognize the WT target site. The plasmid was linearized with restriction enzyme, *Nhe*I, just downstream of the intron and

used as the transcription template. The precursor RNA generated using Ambion's Megascript T7 Kit contained the intron RNA and flanking exons. I performed self-splicing reactions using the precursor RNA and combined the splicing products with purified LtrA protein to reconstitute RNPs, according to established protocols (Matsuura et al. 1997, Saldanha et al. 1999).

To verify that the *in vitro*-prepared RNPs were active, I performed *in vitro* TPRT reactions. The TPRT assay is an indicator of the targeting activity for a given RNP preparation. A schematic diagram of the TPRT method is shown in Figure 2.1a. For the TPRT assay presented in Figure 2.1b, WT RNP particles were incubated with an unlabeled target plasmid, pBRR3-ltrB, which contained a 269-bp sequence corresponding to the Ll.LtrB insertion site from -178 - +91 from the site of intron insertion. Following reverse splicing and endonucleolytic cleavage of the target plasmid, reverse transcription of the intron incorporated [α-^{32}P]-dTTP and other dNTPs, which were present in the reaction mixture.

The products were analyzed by separating them using agarose gel electrophoresis and autoradiography for detection of ^{32}P-labeled plasmid bands that have incorporated nascent cDNAs. Figure 2.1b shows a typical phosphorimage of an agarose gel containing the separated reaction products after incubating the WT RNPs with either the WT target plasmid pBRR3-ltrB (lane 2) or the negative control plasmid pBRR3-CCR5 (lane 1). The annotated autoradiograph shows that the reconstituted RNP particles are capable of carrying out the TPRT reaction and incorporating the radiolabeled-dTTP into high molecular weight plasmid-containing bands. The negative control plasmid confirms that

intron integration occurs preferentially into the pBRR3-ltrB plasmid, presumably at the wild-type target site, while not recognizing an unrelated target site located in the pBRR3-CCR5 plasmid. Signal observed in the autoradiograph below the targeted plasmid corresponds to cDNAs synthesized from free RNAs that are present in the RNP preparation. These RNAs are carried over from the transcription and splicing reactions and are able to act as templates for reverse transcription by LtrA under the TPRT conditions.

2.2 Plasmid Targeting in Zebrafish Embryos

The zebrafish plasmid-targeting assay is based on an assay developed in our lab to test group II intron mobility in *E. coli* (Guo et al. 2000, Karberg et al. 2001). The *E. coli* assay is dependent on the Ll.LtrB intron, which is expressed *in vivo* from the donor plasmid, pACD2, integrating into a target site (the ligated E1-E2 sequence of the *ltrB* gene) cloned upstream of a promoterless *tet*^{R} gene in an Amp^{R} recipient plasmid, pBRR3-ltrB. The intron donor plasmid used for both the zebrafish and the *E. coli* assay contains a modified version of the Ll.LtrB intron in which a phage T7 promoter is inserted in domain IV near the 3' end of the intron replacing the ORF sequence that would normally encode the IEP. Previous work showed that these intron modifications resulted in higher integration efficiencies and increased nuclease-resistance when compared to the full-length intron in the *E. coli* assay (Guo et al. 2001). Insertion of the intron into the target site enables the T7 promoter in DIV to activate the *tet*^{R} gene in the presence of T7 RNA polymerase. Integration efficiencies are measured by determining the ratio of (Tet^{R}+Amp^{R})/Amp^{R} colonies. The *E. coli* plasmid-targeting assay requires delivery of

the donor and recipient plasmid by electroporation to the cells. Expression of the intron and protein from a T7lac promoter in the donor plasmid allows for the formation of active RNPs, which target the recipient plasmid *in vivo*. By contrast, for the zebrafish assay, the WT RNPs are prepared ahead of time and injected into fertilized embryos separately from the recipient plasmid, pBRR3-ltrB.

To assay group II intron targeting reactions in zebrafish embryos, I used WT *in vitro* fertilized zebrafish (*Danio rerio*) embryos. *In vitro* fertilization was performed according to protocols described in *The Zebrafish Book* (Westerfield 2000). The reaction components were injected into the fertilized embryos 15 min after fertilization. A schematic diagram of the protocol used for the plasmid-targeting assay in zebrafish is shown in Figure 2.2. The target plasmid pBRR3-ltrB containing the WT *ltrB* target site was injected first, followed within 1 min by Ll.LtrB RNPs reconstituted, as described above. In all the zebrafish experiments, the DNA target plasmid and RNPs were injected using different needles to avoid prior mixing, and 10 to 25 embryos were injected for each experimental condition. When additional $MgCl_2$ was included in the reaction, it was mixed with the recipient plasmid to reach the desired concentration. After incubation at 30°C for 1 h, the embryos were pooled. Nucleic acids were extracted from the 10-25 pooled embryos and electroporated into *E. coli* HMS174(DE3), which expresses phage T7 RNA polymerase. The cells were then plated on LB medium containing ampicillin with or without tetracycline to determine the mobility efficiency, as described above. Colonies growing on plates containing both Tet^R and Amp^R were picked and used to inoculate liquid LB medium containing ampicillin and tetracycline. The cultures were

incubated overnight at 37ºC while shaking. To verify proper integration of the intron, plasmids were isolated from the overnight cultures and sequenced using a primer, pBRR-MCS, that binds upstream of the target site and generates the 5' junction sequence (Figure 2.3a). Sequence from the isolated plasmids confirmed the intron integration at the target site in the recipient plasmid (Figure 2.3b).

Initial plasmid targeting experiments in zebrafish were performed by injecting Ll.LtrB RNPs at a concentration of 0.1 mg/mL in a solution containing 0.25% phenol red, 10 mM KCl, 10 mM $MgCl_2$ and 40 mM HEPES, pH 8, while the target plasmid was dissolved in 0.25% phenol red in water. These experiments determined that basal targeting was detectable at low efficiency (0.004%), as determined by the colony numbers. However, owing to the previously mentioned studies indicating that group II introns are particularly sensitive to low Mg^{2+} concentrations, we followed-up with experiments in which additional $MgCl_2$ was added to the target plasmid solution at concentrations ranging from 0 to 500 mM. High concentrations of $MgCl_2$ could not be included with the RNP because $MgCl_2$ causes the RNP to aggregate and clog the injection needles. Under these conditions, we measured the group II intron-integration efficiency. This procedure allowed us to determine the optimal injected Mg^{2+} necessary for group II intron integration.

As shown in Figure 2.4a and b, for zebrafish embryos, the highest integration efficiencies were obtained when the injected Mg^{2+} concentration was approximately 150 mM. Given that the zebrafish embryo has a volume of ~200 nL and the injection volume of the DNA/$MgCl_2$ solution was ~10 nL, the calculated increase in intracellular Mg^{2+}

resulting from the injection was 7.5 mM or higher if one corrects for the 77% of the embryo made up of the less permeable yolk (Hagedorn et al. 1998). Therefore, the optimal Mg^{2+} concentration for group II intron targeting in zebrafish embryos is in the range of the optimum Mg^{2+} concentration for the DNA integration reaction (10 mM), as determined by *in vitro* experiments (Saldanha et al. 1999).

While the optimal injected Mg^{2+} concentration was determined to be ~150 mM, when I performed experiments at 125 mM $MgCl_2$, I found that the integration efficiency varied with different batches of RNPs and different batches of embryos such that the maximal targeting efficiencies ranged from 0.11% to 0.89%. Further, in parallel experiments performed in *Xenopus laevis* by others in our lab, it was determined that the addition of dNTPs to the plasmid DNA/$MgCl_2$ solution improved the reproducibility and overall targeting efficiency (Mastroianni et al. 2008). By adding 3.125 mM each of dATP, dCTP, dGTP and dTTP under the pre-determined optimal magnesium conditions and using the highest quality RNPs, I was able to improve the targeting efficiency over 600-fold above the basal targeting efficiency from 0.004% to 2.5%.

Of course, the physiological effects of adding these reaction components to early stage zebrafish embryos needed to be analyzed in order for this to be considered a viable gene targeting method. Accordingly, I examined the impact of the injections on the survivability of the embryos. I found that, for zebrafish embryos, hatch rates after the double injection of the highest concentration of $MgCl_2$ (500 mM) in the DNA solution followed by injection of the RNPs were 26-52% compared to 35% for a single injection of distilled water (Table 2.1). Thus, zebrafish embryos appear reasonably tolerant of the

high intracellular Mg^{2+} concentrations required to support high efficiency group II intron integration. Together, these experiments show that under appropriate conditions, the site-specific group II intron-integration reaction can occur efficiently in zebrafish embryos without significant impact on embryo viability.

2.3 Plasmid Targeting in *Drosophila* Embryos

After demonstrating the activity of group II introns in zebrafish embryos, we set out to determine whether similar results could be obtained in embryos of the fruit fly, *Drosophila melanogaster*. I performed plasmid targeting in *Drosophila* embryos using the same reaction components from the zebrafish experiments. The WT RNPs were prepared, as previously described, using the WT Ll.LtrB-ΔORF intron donor plasmid as the template for generating the intron RNA. Further, the target plasmid, pBRR3-ltrB, was also identical to that used for the zebrafish injections.

The injection techniques differed somewhat from those for zebrafish. Rather than performing *in vitro* fertilization as with zebrafish, *Drosophila* embryos were obtained by setting up mating cages of wild-type (Oregon-R) flies. Shortly after the females laid the fertilized eggs, the embryos were transferred to a block of agar where they were aligned along the edge until they were transferred to a glass cover slip coated with a fine film of glue to hold the embryos. The embryos were briefly desiccated and covered with oil to prevent further drying. Once the embryos were covered with oil, the cover slip holding the embryos was placed on a support block for injection.

For the *Drosophila* injections, the volumes of RNP-containing solution and plasmid DNA solution were based on methods previously established for fly injections in

the laboratory of Dr. David Stein. In this pilot study, the injection volume was not calculated exactly, however it was in the range of 0.5 nL based on visual estimates. A total of 70 embryos were injected with both the RNP and the recipient DNA solutions. Similar to the zebrafish experiments, the injection of RNP was done separately from the DNA injection to prevent mixing prior to entering the embryo. After both injections were complete, embryos were incubated at 30ºC for 30 min. The embryos were subsequently removed from the cover slip by pipetting the entire collection of embryos from the slide and transferring them to an Eppendorf tube containing phenol:chloroform:isoamyl alcohol (25:24:1 by volume; phenol-CIA) for DNA extraction and removal of the oil coating. To increase the volume of the aqueous phase for the phenol-CIA extraction, 50 µL of distilled H_20 was added to the mixture. After the extraction, the aqueous phase, presumably containing the embryo DNA was further diluted by adding 500 µL of SNET buffer and 10 µL of 20 mg/mL proteinase K in water. The solution was incubated for 1 h at 55°C and again extracted with phenol-CIA to remove residual lipids and proteins. The aqueous phase was subjected to ethanol precipitation to purify the nucleic acids. The resulting pellet was resuspended in distilled water for transformation. *E. coli* HMS174(DE3) were electroporated with the extracted nucleic acids and the cells were subsequently plated on LB medium containing ampicillin with or without tetracycline to determine the mobility efficiency, as previously described. Three separate transformations using the same extracted DNA sample showed that without any added $MgCl_2$, the basal targeting efficiency in *Drosophila* embryos ranged from 0.003% to 0.008%. Theoretically, additional $MgCl_2$ in the reactions would have

increased the targeting efficiencies, as in the zebrafish experiments. In fact, subsequent experiments performed by others in our lab verified that targeting efficiencies could be improved with the addition of $MgCl_2$ and dNTPs, increasing targeting efficiencies to as high as 9.5% (Mastroianni et al. 2008).

2.4 Genome Targeting in Zebrafish

Although plasmid targeting in zebrafish embryos was a significant accomplishment and offered the promise of using group II introns in eukaryotic cellular systems, genome targeting with group II introns is the ultimate goal. To this end, we sought to site-specifically insert a group II intron into the zebrafish genome *in vivo*. We chose to target the *mitf* (microphthalmia-associated transcription factor) gene because of its role in the development of the pigmented melanophore cells that give zebrafish their characteristic stripes (Lister et al. 1999). *Mitf* is a basic helix-loop-helix/leucine zipper transcription factor that when inactivated leads to the phenotype known as *nacre* (Lister et al. 1999). Homozygotes for this mutation lack melanophores throughout embryonic and larval development (Figure 2.5a) (Lister et al. 1999). The absence of melanophores provides a quick and easy method of screening embryos for the *nacre* mutation since they lack the distinctive striped skin pattern for which the zebrafish is named.

In order to test the ability of a group II intron to disrupt the *mitf* gene, we had to obtain an intron that successfully targeted the gene sequence. To do this, we used a selection-based method established in our lab for isolating introns that insert into specific sequences (Guo et al. 2000). We cloned the *mitf* cDNA sequence into a recipient vector upstream of a promoterless tet^R gene and selected an intron from a combinatorial library

of introns with randomized target site recognition sequences (EBS and δ) (Guo et al. 2000). We isolated an intron that inserts into the *mitf* gene 235 nucleotides downstream of the start codon (Figure 2.6a). The insertion site is located approximately 24 nucleotides upstream of the activation domain coding sequence (Figure 2.6b). We called this intron mitf235s because it inserted into the sense strand of the *mitf* coding sequence at nucleotide position 235. We called the donor plasmid pACD-mitf235s. In the *E. coli* plasmid-targeting assay, the *mitf*-targeted intron was shown to have a mobility frequency of approximately 52% compared to wild-type (data not shown). When preparing RNPs, we used the linearized donor plasmid pACD-mitf235s as a transcription template to generate the intron RNA, as previously described. *In vitro* targeting activity for the *mitf*-targeted RNPs was confirmed by TPRT (not shown).

To test the *in vivo* activity of the RNP, I performed plasmid-targeting assays in zebrafish embryos using the *mitf*-targeted RNP, as described for the WT RNPs. These experiments were performed before I had determined the optimal plasmid targeting conditions. Thusly, the amount of $MgCl_2$ included in these reactions (20 mM) was significantly below the optimal. Therefore, I was unable to detect plasmid targeting using the *E. coli*-based plasmid assay. Despite suboptimal conditions, I was able to PCR amplify the 5' integration junction from plasmids extracted from the injected embryos (Figure 2.3c and d). Having determined that the mitf-targeted RNPs retained targeting activity *in vivo*, we proceeded with experiments designed to target *mitf* at the genomic level in zebrafish embryos.

For *in vitro* fertilization, we mated wild-type zebrafish with fish homozygous for the *nacre* mutation. This ensured that the F1 progeny were heterozygous at the *nacre* locus. The rationale for using heterozygotes was to ensure that only a single targeting event would be necessary to produce zebrafish embryos with the *nacre* phenotype. Embryos were fertilized *in vitro*, as previously described, and RNPs were injected up to the eight-cell stage. Prior to this developmental stage, the normal permeability barriers between embryonic cells had not yet formed (Kimmel & Law 1985) and persistent bridges between the yolk cell and temporary bridges connecting sibling blastomeres allow for exchange of cytoplasmic contents (Kimmel & Law 1985). Injections following the fifth cell division would prevent uniform distribution of the RNP, due to the formation of the permeability barriers, and the likelihood of mosaic fish would increase. Therefore, I avoided injecting fish after the third cellular division. After injection, embryos were incubated at 28.5°C for 72 h, a period well after the pigmented melanophores become observable (24-48 hours) (Lister et al. 1999, Kimmel & Law 1995). Each injection experiment resulted in numerous embryos, ranging from 50 to 400 in a single injection session, to be screened for pigmentation mutants.

Table 2.1 shows a sample of the targeting experiments and summarizes the results of the screens. Despite testing several different conditions including various dilutions of the RNPs, none of the embryos injected with the *mitf*-targeted RNPs exhibited a phenotype consistent with the *nacre* mutation. In several experiments, unusual phenotypes were observed, but we were unable to associate these deformities with the

group II intron (Figure 2.5b). It is common for embryos to be deformed simply due to injection injury.

A number of limitations regarding the zebrafish injection experiments led us to move away from this line of experimentation. The central reason for ending these experiments was that the number of embryos that could possibly be injected and screened was limited to such an extent that rare targeting events would be difficult to detect. An expert in zebrafish injections might be able to inject as many as 500-1000 embryos in a single experiment. With survivability ranging from 50 to 75%, in a successful injection session, this would require a genome targeting efficiency of at least 1-4 x 10^{-3} for detection of a single targeting event. Although the zebrafish plasmid-targeting assay confirmed that the RNP was capable of inserting into plasmid-borne targets with efficiencies as high as 2.5%, these plasmids were co-injected with the RNPs into the cytoplasm and lacked some of the barriers associated with genome targets such as the nuclear envelope, histones and other chromatin-associated factors. Therefore we had no way of effectively determining the likelihood of isolating a genome-targeting event in zebrafish. I decided that an experimental approach that allowed for larger-scale screens would be more likely to yield the kinds of numbers necessary for isolating the rare events expected from genome targeting with group II intron RNPs in eukaryotes.

2.5 Discussion

In this chapter, I have shown that group II introns retain gene-targeting activity even after injection into live zebrafish and *Drosophila* embryos. This is the first time group II introns have been shown to possess targeting capabilities in these two organisms

and offers a great deal of promise for the application of group II introns as a novel eukaryotic gene targeting technology. While initial plasmid-targeting results revealed limited targeting efficiency for both zebrafish (0.004%) and *Drosophila* (0.008%), these results were consistent with low-level targeting observed for similar experiments previously performed using mammalian cells in our lab (Guo et al. 2000). However, not only was I able to confirm that group II introns are active in zebrafish embryos, I also improved the targeting efficiency as much as 600-fold by injecting $MgCl_2$ along with the target plasmid. Further, recent experiments performed by others in our lab showed that plasmid targeting in *Drosophila* can similarly be enhanced by the addition of $MgCl_2$ from background targeting levels to greater than 9.5% (Mastroianni et al. 2008).

The sensitivity of the group II intron to the concentration of Mg^{2+} in zebrafish embryos is consistent with previous results observed for group II intron splicing activity in yeast mitochondria (Gregan et al. 2001). In yeast mitochondria, decreased Mg^{2+} concentrations strongly inhibit the splicing of four different group II introns belonging to two different subclasses while having little effect on the splicing of any of the group I introns. The impact of Mg^{2+} concentrations on group II intron splicing seems to be due to an effect on the intron RNA since each of the introns affected uses a different protein co-factor (Gregan et al. 2001). While the inhibitory effect of decreased Mg^{2+} in the yeast mitochondria experiments was specific to forward splicing, it can be assumed that reverse splicing, which is necessary for the gene targeting reaction, would be similarly affected. In the event that decreased Mg^{2+} concentrations inhibit both forward and reverse splicing, overcoming this problem must be considered for both reactions when expressing the

group II intron *in vivo* for gene targeting. However, for our experiments, the injection of *in vitro*-prepared RNPs eliminates the need to overcome any inhibitory effect of low magnesium concentration on the forward splicing reaction. Thus, when injecting group II intron RNPs into zebrafish and *Drosophila* embryos, it is only necessary to include additional Mg^{2+} to stimulate the reverse splicing reaction.

With this in mind, for zebrafish, I was able to determine the optimal concentration of Mg^{2+} necessary to enhance the plasmid targeting efficiency. When injecting the RNP at concentrations between 100 and 1000 ng/μL, I determined that highest targeting efficiencies were observed when the plasmid DNA was injected with approximately 150 mM $MgCl_2$. Because the target DNA and $MgCl_2$ were injected in a total volume of 10 nL and the embryo has a volume of ~200 nL, the calculated increase in intracellular concentration of Mg^{2+} (7.5 mM) approached the optimal *in vitro* targeting concentration of 10 mM (Saldanha et al. 1999). It was common to observe targeting efficiencies between 0.18% and 0.89% even when the optimal concentration of $MgCl_2$ was included in the reaction. However, using highest quality *in vitro*-prepared RNPs and the addition of 3.125 mM dNTPs to the target plasmid/$MgCl_2$ mixture led to plasmid-targeting efficiencies up to 2.5% in zebrafish embryos. Targeting efficiencies of this magnitude would be highly desirable for genome targeting so I sought to determine whether group II introns were similarly effective for genome targeting in zebrafish.

Experiments designed to test the usefulness of group II introns for genome targeting in zebrafish relied on the methods developed for plasmid targeting. An intron designed to insert into the zebrafish *mitf* gene was isolated from a combinatorial library

of introns with randomized target site recognition sequences. For *in vitro*-prepared RNPs, the intron RNA was transcribed from the linearized *mitf* donor plasmid. RNPs were reconstituted by combining the excised intron lariat from the self-splicing reaction of the intron RNA with purified LtrA protein. A series of genome targeting experiments was performed, wherein the RNPs were injected into heterozygous *nacre* zebrafish embryos during the one-cell to eight-cell stages, using different concentrations of RNPs with and without a second injection of additional of $MgCl_2$. The expected phenotype consistent with disruption of the *mitf* gene was not observed for any of the targeting experiments. While these results reflect a low gene targeting efficiency, they should in no way be construed to assume that genome targeting is unachievable in zebrafish. Since the time of these experiments, we have significantly improved our ability to produce highly active group II intron RNPs. Further, because these experiments were performed before determining the optimal conditions for plasmid targeting, it can be assumed that experiments performed today could potentially offer a higher likelihood of success. Finally, while supplemental Mg^{2+} is necessary for efficient targeting activity, the addition of Mg^{2+} by injection separate from the RNP is a viable approach for future gene targeting experiments based on survivability results of double injections.

For the experiments described here, we set out to address factors in eukaryotic cells that might inhibit or reduce the gene targeting activity of group II introns. I confirmed that the intracellular magnesium level in zebrafish embryos limits the *in vivo* activity of the RNP. To overcome this impediment, I raised the intracellular magnesium concentration during the targeting reaction by injecting additional magnesium with the

plasmid target with the impressive effect, increasing the targeting efficiency more than 600 fold (from 0.004% to 2.5%) in zebrafish embryos. We also sought to determine whether the intracellular environment would be hostile to the RNP. Our results suggest that any negative cellular response to the intron was not sufficient to abrogate the plasmid targeting activity of the RNP in zebrafish embryos. Further, I also confirmed that plasmid-targeting activity was retained in *Drosophila* embryos. These two metazoan species are sufficiently evolutionarily divergent to suggest group II introns may be active in a broad range of eukaryotes.

Although we wanted to extend the plasmid targeting ability of the group II intron to target the genome, I did not observe any genome targeting events when injecting RNPs targeted to the zebrafish *mitf* gene. This could be due to many complicating factors such as failure of the RNP to localize to the nucleus or inhibition of targeting by the chromatin structure. Even though the genome targeting experiments were inconclusive, it should not be assumed that the obstacles for genome targeting are insurmountable. Future experiments using improved RNP preparations and thoroughly applying some of the lessons learned from the plasmid targeting experiments here may prove that group II introns are capable of targeting the zebrafish genome.

(a) (b)

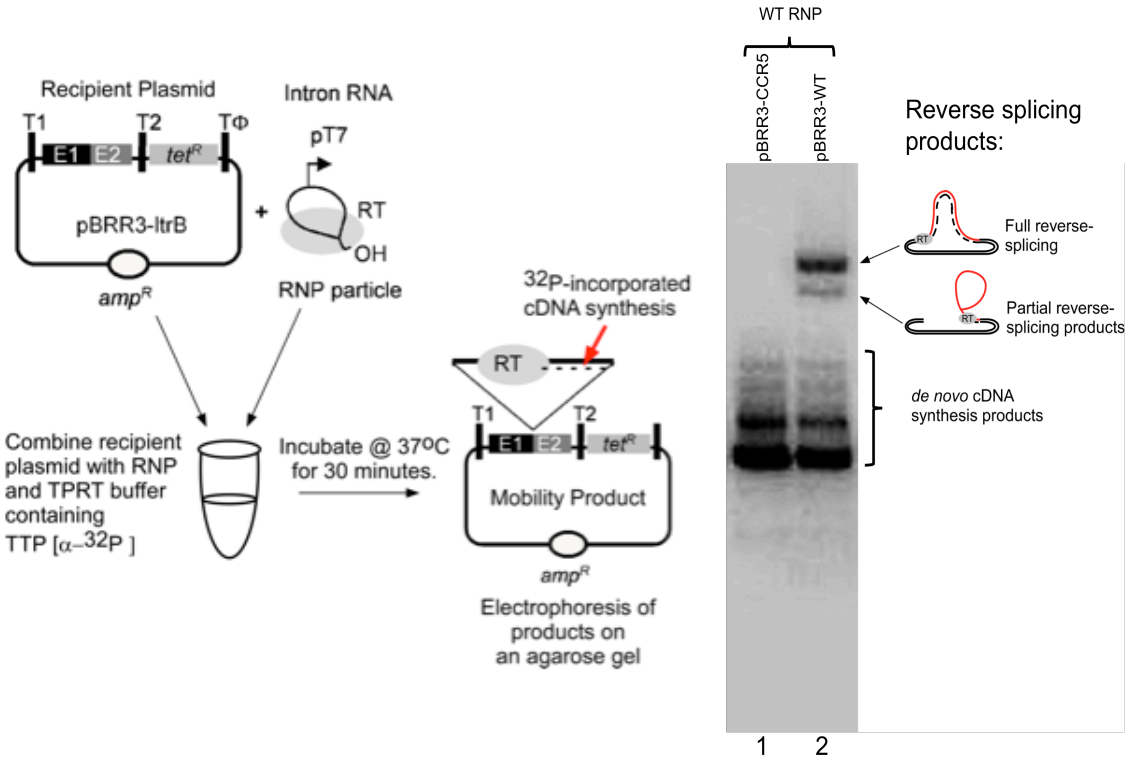


Figure 2.1 Schematic Diagram of the Target Primed Reverse Transcription (TPRT) Reaction and Products Formed When Using WT RNPs.

(a) The TPRT reaction is an assay designed to test the targeting activity of a group II intron RNP *in vitro*. The RNP particle is incubated with an unlabeled target plasmid, which contains a sequence corresponding to the intron-insertion site in 20 μL of TPRT Buffer [10 mM KCl, 10 mM $MgCl_2$, and 50 mM Tris-HCl (pH 7.5)] containing dATP, dGTP, and dCTP (0.2 mM each), and 20 μCi of [α-^{32}P]dTTP (3000 Ci/mmol; New England Nuclear, Boston, MA). Following reverse splicing and endonucleolytic cleavage of the target plasmid, reverse transcription of the intron incorporates the [α-^{32}P]-dTTP and other dNTPs. The reactions are initiated by addition of the RNP particles, and the mixtures are incubated for 30 min at 37°C. Products are analyzed in a 1% agarose gel, which is dried and subsequently autoradiographed with a Typhoon TRIO phosphorimager (Amersham Bioscience). (b) A typical phosphorimage of the agarose-separated TPRT products after incubating the WT RNPs with either pBRR3-ltrB (lane 2) or the negative control plasmid pBRR3-CCR5 (lane 1). The autoradiograph confirms that the reconstituted RNP particles can carry out the TPRT reaction *in vitro*, incorporating the radiolabeled dTTP into high molecular weight plasmid-containing bands. Lane 2 confirms that intron integration occurs preferentially into the pBRR3-ltrB plasmid while not recognizing an unrelated target site located in the pBRR3-CCR5

plasmid (lane 1). Signal observed in both lanes of the autoradiograph below the targeted plasmid corresponds to cDNAs synthesized from free RNAs that are present in the RNP preparation. These RNAs are carried over from the transcription and splicing reactions and are able to act as templates for reverse transcription by LtrA under the TPRT conditions.

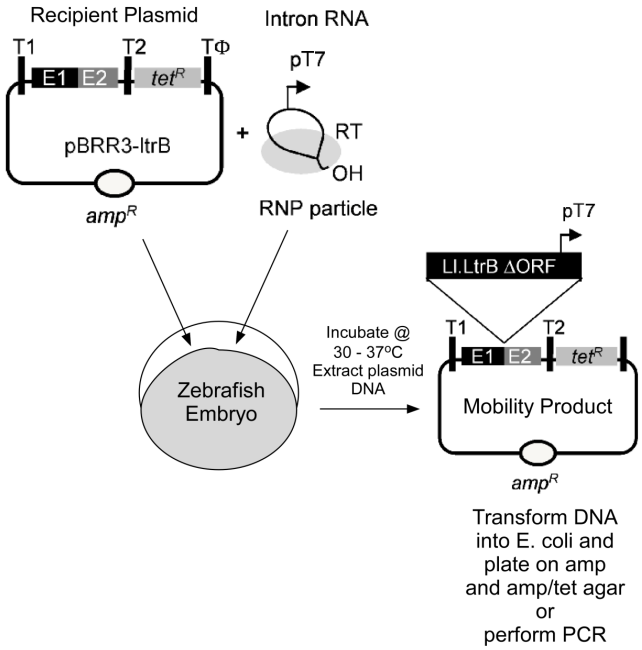


Figure 2.2 Schematic Diagram of the Group II Intron Plasmid-Targeting Assay in Zebrafish Embryos.

The *in vivo* plasmid-targeting assay involves separate injections of the WT group II intron RNP and the target plasmid with specified amounts of $MgCl_2$ into early stage zebrafish embryos. The Ll.LtrB-ΔORF intron has a T7 promoter cloned into DIV. Integration of the intron into the target site cloned upstream of a promoterless *tet*^{R} gene in an Amp^R target plasmid (pBRR3-ltrB) activates the expression of the *tet*^{R} gene. After the embryos are injected and incubated for 30 min at 30°C, nucleic acids were isolated and electroporated into *E. coli* HMS174(DE3). The integration efficiency is calculated as the ratio of (Tet^R+Amp^R)/Amp^R colonies.

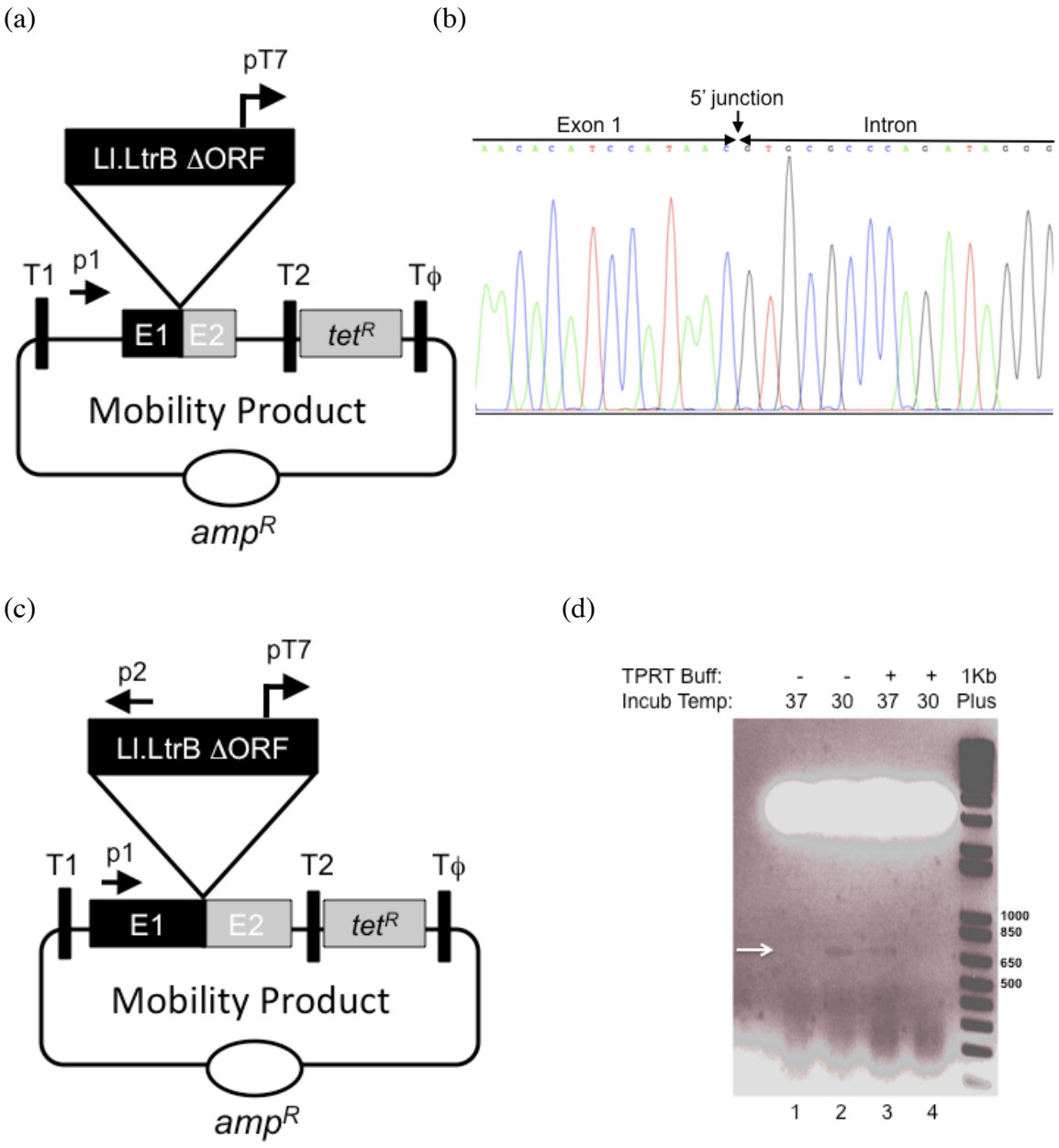


Figure 2.3 Sequencing and PCR to Detect Site-Specific Intron Integration for Plasmid-Targeting Assays in Zebrafish Embryos.

(a) Verification of intron integration into the WT target site was confirmed by sequencing the target site of the target plasmid following extraction of nucleic acids from injected zebrafish embryos. The zebrafish embryos were injected with WT RNPs separately from the recipient plasmid, pBRR3-ltrB to avoid prior mixing. 10 to 25 embryos were injected for each experimental condition and incubated at 30°C for 1 h.

After incubation, the embryos were pooled and nucleic acids were extracted and electroporated into *E. coli* HMS174(DE3). The cells were then plated on LB medium containing ampicillin with or without tetracycline to determine the mobility efficiency based on the ratio of (Tet^R+Amp^R)/Amp^R colonies. Colonies growing on plates containing both Tet^R and Amp^R were picked and used to inoculate liquid LB medium containing ampicillin and tetracycline. The cultures were incubated overnight at 37ºC while shaking. To verify proper integration of the intron, plasmids were isolated from the overnight cultures and sequenced using a primer, pBRR-MCS (p1), which binds upstream of the target site and generates the 5'-junction sequence. (b) Sequence from the isolated plasmids confirmed the intron integration at the target site in the recipient plasmid. (c) Similar to the assay for WT plasmid targeting, *mitf*-targeted RNPs were injected into zebrafish embryos separately from the target plasmid, pBRR3-mitf. Following incubation, nucleic acids were extracted and used as a template for PCR to detect intron integration. Primers used to detect intron integration were LtrB reverse bottom 1 and Upstream mitfa/b Top (1). (d) The expected size of the PCR product corresponding to the 5' integration junction for the *mitf*-targeted RNP in pBRR3-mitf was 582 bp.

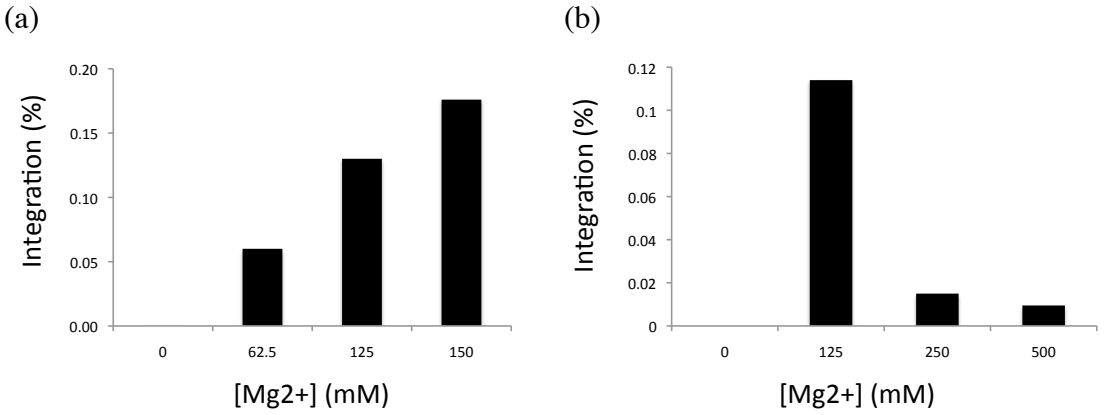
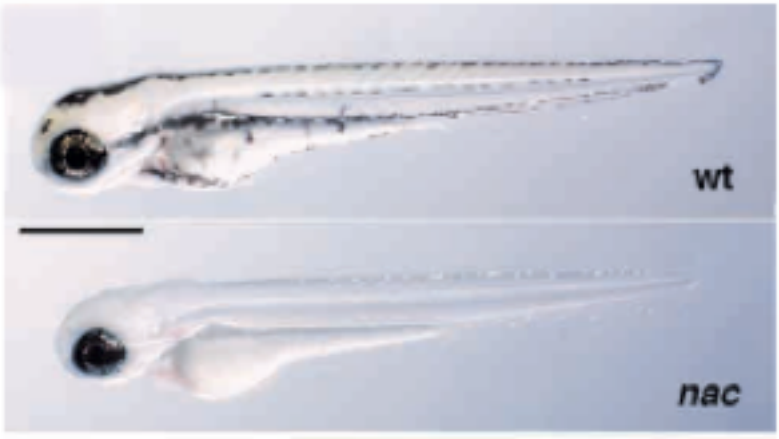


Figure 2.4. Determination of Optimal Magnesium Concentrations for Site-Specific Integration of Group II Intron RNPs into a Plasmid Target Site in Zebrafish Embryos.

For group II intron plasmid targeting assays in zebrafish embryos, target plasmid (0.5 mg/ml in a solution containing 3.125 mM of each dNTP and specified amounts of $MgCl_2$, with 0.25% phenol red) and Ll.LtrB-ΔORF RNPs (0.1-1.0 mg/mL with 0.25% phenol red, 10 mM KCl, 10 mM $MgCl_2$, and 40 mM HEPES, pH 8.0) were injected separately into 10 to 25 one-cell to eight-cell embryos. The injection volumes were ~5-10 nL. After injection, the embryos were washed with Steinberg's medium, pooled in a single 1.5-ml Eppendorf tube in 500 µl Steinberg's medium, and incubated for 30 min at 30°C, prior to nucleic acid extraction and transformation of *E. coli* HMS174(DE3). (a) Zebrafish embryos were injected with target plasmid DNA solution containing 0 to 150 mM $MgCl_2$ and (b) 0 to 500 mM $MgCl_2$ followed by injection of WT Ll.LtrB RNPs. Integration efficiencies (%) for both experiments were determined based on the ratio of $(Tet^R+Amp^R)/Amp^R$ colonies. Optimal plasmid targeting was observed when the target plasmid was injected with 125-150 mM $MgCl_2$.

(a)



(b)

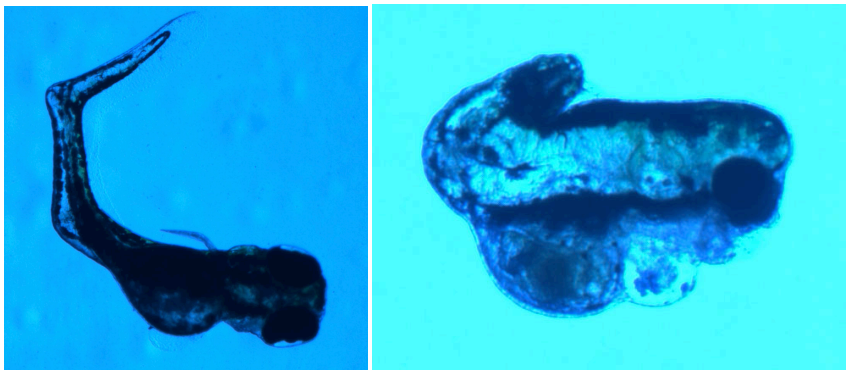
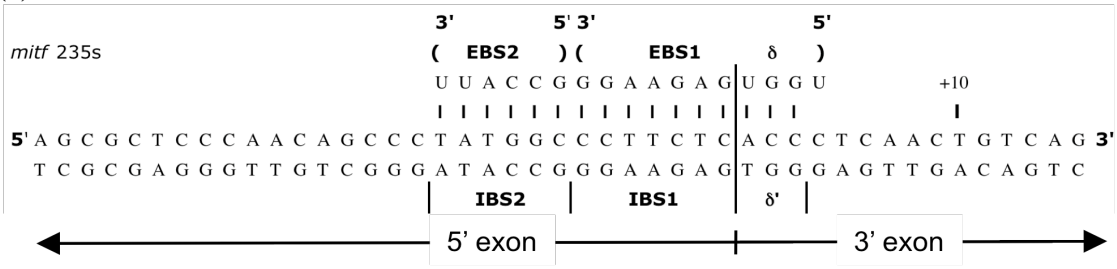


Figure 2.5 The *nacre* Mutant and Deformed Embryos Observed During *mitf* Targeting Experiments

(a) Lateral view of the wild-type (top) and *nacre* (bottom) larvae at 3-days post-ferilization. Embryos homozygous for *nacre* lack all neural crest-derived melanophores. Adapted from Lister et al., 1999. (b) "Bent Tail" and "Anteriorized" mutants observed during injections targeting the *mitf* gene in zebrafish embryos. Similar deformities were commonly observed in the genome targeting experiments, but could never be associated with activity of the RNP.

(a)



(b)

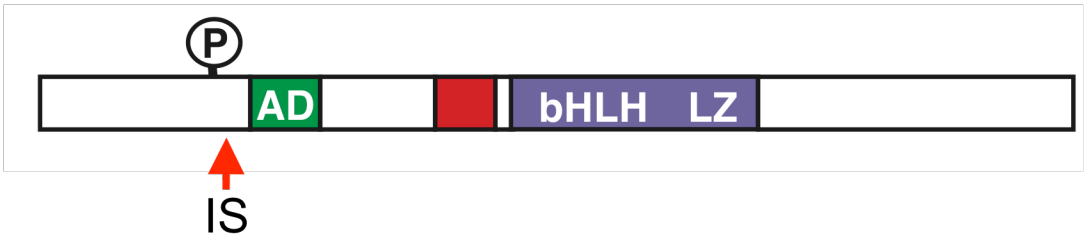


Figure 2.6 Ll.LtrB Intron Target Site for the *mitf* Gene in the Zebrafish Genome and the Corresponding Location in the Mitf Protein

(a) Alignment of the intron EBS1, EBS2 and δ sequences with the DNA target site sequences (IBS1, IBS2 and δ') in the zebrafish genomic sequence for the *mitf* gene. The *mitf*-targeted intron was selected from a combinatorial library of introns with randomized target site recognition sequences (EBS and δ) (Guo, 2001). The intron inserts into the *mitf* gene 235 nucleotides downstream of the start codon. (b) The insertion site (red arrow) is located approximately 24 nucleotides upstream of the sequence encoding the activation domain of the Mitf protein.

Table 2.1 Zebrafish Genome Targeting.

Injection Stage	[RNP] (μg/mL)	[$MgCl_2$] (M)	Survival	Percent Survival	Percent *nacre*
1-cell	Sham injection	0.00	29/50	58%	0%
1-cell	DNA in water	0.00	40-60/400	10-15%	0%
1-cell	1.0	0.00	8/100	8%	0%
1-cell	1.0	0.00	8/150	5.30%	0%
1-cell	0.1	0.00	23/200	11.50%	0%
1-cell	0.1	0.00	0/50	0%	0%
8-cell	0.1	0.00	9/50	18%	0%
2-cell	0.1	0.00	20/150	13%	0%
1-cell	0.05	0.00	6/40	15%	0%
1-2 cell	0.15	0.00	14/88	16%	0%
1-2 cell	0.15	0.50	31/60	52%	0%
1-2 cell	0.15	0.50	18/68	26.50%	0%
1-2 cell	dH_2O	0.00	54/156	34.60%	0%

The results of a series of genome targeting experiments are summarized in this table. Despite injections at various stages with different amounts of RNP and in some cases high concentrations (500 mM) of $MgCl_2$, no embryos were observed with the *nacre* phenotype. Importantly, the viability of embryos was not affected by injections with 500 mM $MgCl_2$.

CHAPTER 3: GROUP II INTRON-BASED GENE TARGETING IN MAMMALIAN CELL CULTURE

Because the original experiments in our lab demonstrated that group II introns retain retrohoming activity for plasmid targets in eukaryotic cells, a great deal of research has been carried out to direct them to target the chromosomes (Guo et al. 2000, Mastroianni et al. 2008). Unpublished experiments performed in our lab by others have produced additional data to suggest that it is only a matter of time before group II introns will become a breakthrough technology for eukaryotic gene targeting. For example, it has been verified by PCR that group II introns can locate and insert into specific target sites even within the highly complex environment of purified chromosomal DNA via *in vitro* TPRT reactions (Cui 2005). Group II introns have also targeted chromosomal DNA in mammalian cell nuclear lysate without additional Mg^{2+} suggesting that intracellular Mg^{2+} levels are sufficient for gene targeting (Cui 2005). In addition, Ll.LtrB group II intron RNPs have been delivered to mammalian cells by electroporation and immunofluorescence using an anti-LtrA antibody showed that at the protein component of the RNP localized to the nucleus of HEK 293 cells. Nuclear localization of the protein does not guarantee that active RNPs also did so. However, it was confirmed by PCR that electroporated group II intron RNPs designed to insert into multicopy rDNA targets site-specifically integrated into the chromosome (Cui 2005). This suggests that some active RNPs had properly localized to the nucleus. Despite successful PCR amplification of these targeting events, isolating a viable mammalian cell with a group II intron-based targeting event has yet to be accomplished. In addition, insertion events to date have

only been detected when targeting multicopy target sites or a single copy reporter construct. Targeting of a single-copy endogenous gene has not yet been shown.

Developing the best method for using group II introns in mammalian cells is critical for this technology to become useful for gene therapy-related purposes. In pursuit of this method, we have tested group II introns for eukaryotic gene targeting in a variety of experiments. Some of the drawbacks of previous experiments performed in our lab include: 1) the lack of a positive control for experimental comparisons; 2) the inefficiencies associated with the group II intron targeting activity in eukaryotic cells, especially regarding single target sites; and 3) the inability to enrich for and isolate living cells containing group II intron-based targeting events. Developing an assay that addresses these problems was a goal that I chose to pursue. Meeting this goal would improve the overall experimental approach for understanding the targeting behaviors of and the conditions necessary for group II introns in eukaryotic cells.

By combining the unpublished data from our lab with existing gene targeting methods, I developed the experimental approaches described in this chapter. I will discuss experiments in which I use group II intron RNPs to create a double-strand break at a single target site in the genome that facilitates, via intermolecular recombination, the site-specific incorporation of an exogenous donor DNA molecule. While the first demonstration of RNP-stimulated integration used a linearized double-stranded donor DNA, use of the single-stranded genome of recombinant adeno-associated virus (rAAV) improved our method of donor DNA delivery. The rAAV experiments generated promising data to suggest that group II introns can stimulate integration of the donor

DNA at the target site and act as a eukaryotic gene targeting technology. These experiments offer at least two improvements upon the experimental design of previous gene targeting experiments in our lab. First, we can isolate cells containing site-specific targeting events that have presumably been stimulated by group II intron induced double-strand breaks. Second, by using rAAV to deliver the donor DNA, we have a reliable method for determining a positive control level of targeting that serves as the baseline for comparisons with group II intron experiments. Through these improved experimental approaches, I have obtained data to suggest that group II intron RNPs can stimulate the site-specific integration of exogenous DNAs.

3.1 Preparation of CCR5-Targeted Group II Intron RNPs and *In Vitro* Targeting Experiments

In 2001, Guo et al. designed group II introns to insert into therapeutically relevant targets. The β-chemokine receptor gene, CCR5, was chosen as a potential target because of its well-understood role in facilitating HIV infection of T lymphocytes (Deng et al. 1996). CCR5 is commonly used for novel gene targeting technologies as a proof-of-principle target gene. Several introns were designed to target the CCR5 gene with varying degrees of targeting efficiency (Figure 3.1b). Of the five group II introns designed to target CCR5, one had a particularly high targeting efficiency of 53% +/- 8% compared to the WT Ll.LtrB intron in the *E. coli* mobility assay (Figure 3.1b) (Guo et al. 2000). The insertion site for this intron was located at nucleotide 332s of the major product of the CCR5 cDNA, which at the genomic level corresponds to exon 4 (Figure 3.1a and c) (Mummidi et al. 1997). Guo et al. prepared purified RNPs using the CCR5-

targeted intron. HEK 293 cells were transfected with the RNPs and a plasmid DNA containing the CCR5 target site using a liposomal formulation (DMREI-C) designed for efficient nucleic acid transfection of mammalian cells (Guo et al. 2000). Nested PCR confirmed that the RNPs located and inserted into their target sites in the plasmid DNA. These experiments are the basis upon which I designed new experiments for targeting the CCR5 gene in human cells using the CCR5-targeted RNP.

In order to utilize CCR5-targeted RNPs for genome targeting, it was necessary to produce high quality *in vitro*-prepared RNPs. Following RNP preparation procedures similar to those described in Saldanha et al. (1999) and previously described for the zebrafish RNP injection experiments, I prepared CCR5-targeted RNPs for mammalian genome targeting experiments. The RNP particles were reconstituted by combining purified LtrA protein and excised lariat intron RNA obtained by *in vitro* self-splicing of the CCR5-targeted Ll.LtrB RNA (pACD2-CCR5(332s)/*Nhe*I *in vitro* transcript). As previously mentioned, the RNA included in the RNP reconstitution contains heterogeneous RNAs 60-70% of which is intron lariat while the remainder consists of precursor RNA and splicing products including ligated exons and linear intron RNA. Therefore, when determining the RNA concentration by $O.D._{260}$, it is difficult to specify the exact amount of protein to include in the RNP reconstitution. In order to determine the optimal amount of protein to include in the reconstitution reaction, I reconstituted RNPs with different amounts of protein and tested each of the resulting RNPs for TPRT activity (Figure 3.2). The most active RNP, as determined by this method, is used in targeting experiments. The CCR5-targeted RNPs were regularly more active in the

TPRT reactions than WT RNPs prepared in parallel using similar reagents, as can be seen in Figure 3.2.

In addition to the standard plasmid-based TPRT reactions, I performed "cold" TPRT reactions using genomic DNA as the target. Cold TPRT reactions consist of similar reactants to the standard TPRT reaction, but no radiolabeled nucleotides are included in the reaction and purified human genomic DNA replaces the plasmid DNA as the target. In order to detect targeting, I performed PCR to amplify the 3'-integration junction. Successful amplification across the 3' junction suggests that the intron has correctly inserted at its target site and reverse transcription of the intron is complete, up to the annealing point for the PCR primer (Figure 3.3a). The 586-bp PCR product corresponding to the 3' junction was readily amplified from a plasmid targeting experiment that was performed parallel to the genome targeting experiment (Figure 3.3b). Targeting products were less efficiently amplified from the genomic DNA targeting experiment. However, the 492-bp amplicon corresponding to the 3' junction was detectable, verifying that the CCR5 RNP could find its target site even within a more complex genetic environment. Dilution of the genomic DNA in the reactions determined the minimum level of detection under these conditions (Figure 3.3b). Based on the amounts of genomic DNA used in the cold TPRT reactions (50 ng, 5.0 ng and 0.5 ng), I determined that *in vitro* targeting of the CCR5 gene could be visibly detected on an agarose gel in as few as 1,510 genomes (Figure 3.3b), using 3.3 pg as the approximate mass of one human genome. PCR confirmation of *in vitro* targeting suggested that genome integrations would be detectable even among more complex DNA mixtures.

3.2 Delivery of RNPs by Electroporation and *In Vivo* Targeting of CCR5 by Direct Integration

Electroporation uses a high voltage pulse to increase the permeability of the cell membrane in order to increase the transmembrane transport of molecules that otherwise cannot enter the cell. Electroporation is commonly used to deliver DNA and naked RNA to mammalian cells. A less common application of electroporation is for the transfection of cells with ribonucleoprotein complexes. Li et al. developed one of the first RNP electroporation methods using influenza virus RNPs (Li et al. 1995). Subsequently, our lab established appropriate conditions for delivering group II intron RNPs to HEK 293 cells (230 V, 250 μF) (Cui 2005). By fusing GFP to the protein component of the RNP, the protein was shown to localize to the nucleus of the cells (Cui 2005). In follow-up experiments using the established electroporation methods, rDNA-targeted RNPs were delivered to HEK 293 cells. PCR amplified the intron inserted at the appropriate rDNA target site.

For the experiments described in this chapter, I used a similar approach to deliver CCR5-targeted RNPs to two different mammalian cell lines: K562 (human hematopoietic cell line) and HCT116 (human colorectal cancer cell line). The electroporation conditions for these experiments are described in Table 3.1. The electroporation conditions for K562 cells were based on conditions previously shown to deliver DNA using the Gene Pulser II electroporation system (Bio-Rad Laboratories, Electroprotocols Bulletin #D035551). Based on these conditions, the voltage and capacitance were set to 320 V and 800 μF, respectively. Because there was no established protocol for

delivering RNPs to HCT116 cells, I tested both the previously described sets of electroporation conditions from the HEK 293 cell and K562 cell experiments.

To test the conditions for delivering RNPs to K562 cells, I sought to disrupt the CCR5 gene by direct intron integration. Direct integration of a group II intron at the target site is one of the most straightforward ways to disrupt a gene at the genomic level. Although Cui was able to detect targeting events for rDNA after electroporating RNPs into HEK 293 cells, rDNA is a multicopy target. Multicopy targets offer more opportunity for targeting; and their distribution along the chromosome very likely corresponds to varying degrees of chromatin condensation. rDNA genes have a copy number of 400 per human cell and are actively transcribed in dividing cells (Yoon et al. 1995, Gencheva et al. 1996). A group II intron has also targeted a single copy marker gene integrated into the genome, however it is possible that the target integrated at a recombination "hot spot" so this group II intron integration event is not typical of an endogenous gene. Targets like the actively transcribed rDNAs and the integrated selection marker will likely be more targetable due to sufficient chromatin "openness" than more obstructed target sites such as non-transcribed genes. The single copy CCR5 gene is likely to be a more difficult target to disrupt depending on the chromatin condensation around the target site. In order to increase our chances of targeting CCR5, we chose the K562 cell line that is known to express the CCR5 gene (Mummidi et al. 1997, Mondal et al. 2005). Transcription would likely provide a more open chromatin structure. This we hypothesized would improve the accessibility of the target site to the group II intron RNP.

To test our hypothesis, I delivered *in vitro*-prepared CCR5-targeted RNPs by electroporation to mammalian cells in order to disrupt the CCR5 gene by direct integration. Experiments in zebrafish, *Drosophila* and *Xenopus laevis* that demonstrated the enhancing effects of $MgCl_2$ on group II intron targeting activity led me to include $MgCl_2$ in some reactions (Mastroianni et al. 2008). Figure 3.4 shows the results of nested PCR to detect the 3' junction of the intron at the target site following electroporation of the RNP into K562 cells. The same primers used for the cold TPRT experiments performed with purified genomic DNA were used for the primary PCR. The primers used for nested PCR were the intron-specific primer, ltrB 890s, and the CCR5-specific primer, CCR5 +156-+135. In this experiment, while most samples received a total of 10 μg of the CCR5-targeted RNP, only the sample that received 20 μg of RNP had a distinct PCR product of the appropriate size (211 bp) (lane 6, Figure 3.4). The PCR product was not detected in cells receiving no electroporation pulse and/or no RNP. The addition of $MgCl_2$ with 10 μg of RNP did not seem to have any effect on the reaction nor did decreasing the number of cells in the reaction from 5×10^6 to 5×10^5. While these results were promising, I was unable sequence the PCR product to confirm that it reflected a true integration event. Further, there was no way to enrich for cells with the CCR5 gene disruption. This limitation led us to pursue experiments to achieve this important qualification for gene targeting.

3.3 Targeting the CCR5 Gene *In Vivo* by Group II Intron RNP-Stimulated Homologous Recombination

Based on the promising results for the *in vivo* targeting of the CCR5 gene in K562 cells, we felt that the group II intron RNP could potentially be used in at least 3 different ways for gene targeting; 1) to disrupt genes by direct integration, 2) to generate a double-strand break that leads to mutagenic repair by NHEJ, and 3) to generate a double-strand break that stimulates the incorporation of donor DNAs by homologous recombination. In most cases, disrupting genes by direct integration does not allow for enrichment of successfully targeted cells because there is often no selectable phenotype for the disrupted target gene, as is true for the CCR5 gene. Although a selectable marker can be incorporated into the intron, its presence is expected to decrease the efficiency of the group II intron integration reaction (Frazier et al. 2003, Plante & Cousineau 2006). Therefore, use of the group II intron to create a double-strand break that stimulates the integration of an exogenous DNA fragment carrying a selectable marker is an attractive method for targeting and isolating live cells containing gene disruptions.

3.3.1 Designing the Plasmid Donor DNA for Targeting the CCR5 Gene

Our lab has previously demonstrated that lariat-containing group II intron RNPs can create double-strand breaks that stimulate homologous recombination in *E. coli* and in *Xenopus laevis* oocytes (Mastroianni et al. 2008, Karberg et al. 2001). The early steps of the retrohoming process cleave the DNA. First, the 3' end of the intron RNA reverse splices into the top strand of the DNA target site. Next, the IEP cleaves the bottom strand approximately 10 nucleotides downstream of the RNA insertion site. The double-strand

break persists until the 5' end of the intron RNA forms a covalent bond with the 3' end of the 5' exon. As previously depicted in Figure 2.1b, partial reverse splicing, in which the 5' end of the intron fails to bond with the 5' exon, represents a significant percentage of the targeting events in the *in vitro* TPRT reaction and varies with RNP batches (Zhuang et al. 2009, Saldanha et al. 1999). The mechanism by which group II introns stimulate HR-based incorporation of exogenous DNAs is assumed to be similar to that described for other site-specific endonucleases. While both linear and lariat-containing group II intron RNPs can stimulate HR, Mastroianni et al. (2008) showed that group II intron RNPs containing lariat intron RNA have a somewhat higher stimulatory effect than group II introns containing linear intron RNA. Using group II introns as a site-specific endonuclease to facilitate successful integration of the donor DNA will result in disruption of the target gene and will allow for enrichment using the selectable marker present in the donor DNA.

To test whether group II introns can create double-strand breaks that stimulate integration of a donor DNA molecule at the target site in cultured mammalian cells, we designed a pick-up fragment (PUF) donor DNA, pCCR5-PUF2, that would incorporate at the CCR5 gene by homologous recombination (HR) (Figure 3.6a). The design of the PUF construct is based on standard HR DNA constructs used in classical gene targeting experiments in which regions of sequence homology flanking the desired insertion site are located on either side of a selection marker (Doetschman et al. 1988, Bronson & Smithies 1994). Two regions of homology were amplified by PCR from the CCR5 gene; the first beginning 25 bp 5' of the target site and extending 2.2 kb upstream, including a

portion of intron 2 and exon 4, and a second fragment beginning 15 bp 3' of the target site and extending 4.1 kb downstream, consisting of a portion of exon 4 and intron 3 (Figure 3.6a). The 40 base pairs directly flanking the target site were omitted from the construct in order to prevent the RNP from recognizing and targeting the donor DNA molecule prior to or following insertion.

Sedivy and Dutriaux (1999) demonstrated that configuring a targeting construct as a promoter trap such that marker gene expression is dependent on the endogenous gene promoter results in significant enrichment for homologous recombinants within the pool of transgenic clones. Thus, an ATG (-) blasticidin resistance gene ORF was cloned in frame with the homology arm containing the upstream coding sequence of exon 4 from the CCR5 gene. By fusing the blasticidin gene to the CCR5 ORF, expression would necessarily be driven by endogenous transcription from the CCR5 promoter upon integration and successful targeting events could be selected based on blasticidin resistance.

The donor DNA was constructed in a pUC19 plasmid backbone in such a way that linearization (*Mlu*I digestion) of the vector was necessary to free the ends thereby allowing the targeting vector to take on an "ends-out" arrangement when aligned along the targeted chromosome (Figure 3.6a). The "ends-out" design of the targeting vector is commonly considered a "replacement vector" and is often used in yeast and mouse for gene targeting (Figure 3.5) (Müller 1999, Gong & Golic 2003). Essentially, the replacement vector is designed such that homologous sequences within the PUF are colinear with the target sequences. A double crossover event replaces chromosomal

sequences with the vector sequences (Müller 1999). In this case, the blasticidin selection cassette and the pUC19 vector backbone would replace the 40-bp sequence corresponding to the group II intron recognition site (Figure 3.6a). Because the construct was designed to encode a CCR5 truncated fusion protein, cells exhibiting blasticidin resistance due to non-specific integration events would be reduced to only those insertions that resulted in fusions with upstream exons of other genes. We felt this would give us high confidence in any blasticidin resistant colonies that arose during the selection process. By introducing the PUF donor DNA into K562 cells along with the CCR5-targeted RNP, we expected an increased efficiency of PUF integrations at the target site following cleavage of the target DNA as has been observed with other double-strand break-stimulated gene targeting experiments.

3.3.2 Electroporation of K562 Cells with RNP and Donor DNA, Selection Methods and Panning for PUF Insertions

Using similar conditions to those described above for the direct integration experiments, I delivered RNPs along with the linearized pCCR5-PUF2 donor DNA to K562 cells. For this experiments, I referred to electroprotocols established for the Bio-Rad Gene Pulser for the delivery of DNA to K562 cells (Bio-Rad Laboratories, Electroprotocols Bulletin #D035551). Both the donor DNA and the RNP were resuspended in HKM buffer [10 mM KCl, 10 mM $MgCl_2$, and 40 mM HEPES, pH 8.0], which was developed in our lab for optimal resuspension of group II intron RNPs. During the electroporation, cells were suspended at 5 x 10^6 cells/mL in 800 mL of IMDM medium (Invitrogen). This experiment included a constant concentration of donor DNA

(approximately 10 μg) combined with varying concentrations of CCR5-targeted RNP (12.5 μg, 25 μg and 50 μg). The RNP and DNA were added separately to the cell suspension immediately prior to electroporation. Following electroporation (Voltage: 320 V, Capacitance: 800 μF), the cells were transferred to a T-75 culture flask containing 50 mL of growth medium without FBS and incubated at 37ºC for 1 h. After 1 h, the medium was exchanged, and the cells were incubated overnight in complete medium without selection. Although the cell viability was significantly reduced for each of the samples that received the RNP along with the DNA, after 48 hours of recovery, the selection process for gene targeting events was initiated by adding blasticidin (5 μg/mL) to the medium.

Cells resistant to blasticidin were tracked by counting live cells in each of the samples (Figure 3.8). Five days post-electroporation, the concentration of blasticidin was increased to 10 μg/mL. Despite the addition of 10 μg/mL of blasticidin to the medium, the cells maintained a steady increase until day 8, when it appeared that the blasticidin inhibition began to take effect by significantly decreasing the number of live cells. Either by error or technical failure, most samples appear to have been released from selective pressure from day 8 until day 9. However, by day 11, after refreshing the selection medium, the cells again underwent significant cell death. From day 13 until day 21 post-electroporation, the number of live cells in all samples except the positive control (sample 7) was greatly reduced. The positive control sample had been transfected with a plasmid expressing the blasticidin resistance gene from a constitutive promoter and therefore was expected to contain cells resistant to the antibiotic. Beginning on day 14, the positive

control (Sample 7) had nearly 10 times more live cells than each of the other samples. However, by day 26, samples that received both the RNP and the donor DNA (Samples 4-6) began to approach the number of live cells observed in the positive control. By day 29, cells in the sample that received neither the donor DNA nor the RNP (Sample 1) and cells that received either the donor DNA alone (Sample 2) or the RNP alone (Sample 3) were entirely eliminated by blasticidin selection (Figure 3.8). At the same time, each of the samples that received both the RNP and the donor DNA (Sample 4-6) reached comparable cell survival numbers to those of the positive control (Sample 7) (Figure 3.8). At this point, the cell numbers were sufficient to begin DNA analysis to determine whether samples 4-6 contained PUF insertions at the CCR5 target site. DNA was extracted from samples 4-7 to check for integration of the PUF at the target site (Figure 3.8), however the cells at this stage consist of a heterogeneous pool of all the possible integration events. Isolating individual blasticidin resistant cells from this pool will allow for the cell-specific characterization of the integration events, whereas analysis of the genomic DNA extracted at this stage will give an indication of whether site-specific integration has occurred in the heterogeneous pool of cells.

Because the K562 cell line is a suspension cell line, isolating single drug-resistant cells is more difficult than when using an adherent cell line. However, Anderson et al. developed a simple and efficient panning method for separating live cells growing in suspension from dead ones (Anderson & Junker 1994). The method takes advantage of the nearly universal presence of α-glucosyl and α-mannosyl residues on the outer surface of mammalian cell membranes. The method relies on the fact that live mammalian cells

will adhere to plastic or glass surfaces coated with the plant lectin concanavalin A (con A) while dead cells will not. A schematic diagram of the targeting and selection methods is included in Figure 3.7. Each of the samples containing live cells (samples 4-7) was screened using the methods described in Anderson & Junker 1994. Individual cells were picked from the plates where live cells had attached to the conA-treated surface. The cells were transferred to 96-well plates for population expansion in order to generate enough cells for large-scale DNA extractions. DNA extractions were performed on the isolated colonies and PCR was performed to screen for PUF integrations, but none of the selected colonies were conclusively shown to contain a site-specific integration of the PUF.

3.3.3 PCR Analysis of the CCR5 Gene in Cells Targeted by RNP-Stimulated Homologous Recombination Using pCCR5-PUF2

In order to characterize the blasticidin resistant cell populations from the CCR5 targeting experiments, I analyzed the CCR5 gene region using a series of PCR strategies (Figure 3.9a). Initially, the most direct way to detect an insertion at the target site was to amplify the blasticidin ORF at the CCR5 locus. Because the homology arm upstream of the target site was approximately 2.2 kilobases, I designed two primers that annealed 5' of this homology region so as to avoid amplification of the homology region from the donor DNA, p1 (CCR5 exon 1) and p2 (CCR5 exon 1b) (Figure 3.9). Similarly, I designed two primers within the blasticidin resistance gene directed anti-parallel to the coding sequence, p9 (Blasticidin 2 anti) and p10 (Blasticidin 1 anti) (Figure 3.9). The two sets of primers allowed for nested PCR reactions. Using the outermost set of primers

(p1 and p10) including the 5'-most CCR5 primer and one inside the blasticidin ORF for the primary PCR resulted in an inconclusive gel with no specific amplification in any of the lanes corresponding to samples 4-7 or in an additional lane containing PCR products from genomic DNA of untreated K562 cells (sample 8) (Figure 3.9b). Use of the primary PCR products as the template for nested PCR using primers that anneal just inside those of the primary PCR reaction resulted in amplification of a non-specific band in samples 4, 5, 6 and 8 with a streak of non-specific amplification in the lane containing the sample 6 PCR products (lane 3, Nested PCR gel, Figure 3.9b). Upon closer inspection, it appeared that a faint band corresponding to a 2.6-kb PCR product was present in the lane for sample 6 (lane 3, Nested PCR gel, Figure 3.9). Since this was the expected size for the desired amplification product in the event of PUF integration, I analyzed sample 6 genomic DNA with much greater scrutiny.

I designed primers positioned within the CCR5 gene at successively increasing distances upstream of the integration site for the blasticidin ORF (Figures 3.9a). When combined with an anti-parallel blasticidin resistance gene primer (p10, Blasticidin 1 anti), the new CCR5 primers were designed to amplify PCR products ranging from 503 bp to 2,601 bp (Figure 3.9a and c). While each of the smaller PCR products could be amplified from randomly integrated donor DNA, the 2.6-kb fragment could only be amplified if the donor DNA had integrated at the target site (lane 2, Figure 3.9c). The results of this PCR characterization of sample 6 genomic DNA increased our confidence that we had successfully inserted the donor DNA into the CCR5 gene. Surprisingly, the amplification of the PUF insertion always required nested PCR, suggesting that the targeted DNA was

present in very low amounts. Perhaps this is due to the fact that sample 6 was not a homogenous suspension of cells, but rather was a collection of cells many of which did not contain the site-specific PUF insertion. In order to generate enough PCR products for sequencing, it was necessary to perform nested PCR using the 2.6-kb fragment as the template. The nested PCR product as shown in Figure 3.9d was 2,479 bp and was successfully sequenced to confirm the amplification of the donor DNA at the CCR5 target site.

Additional PCR were performed to rule out the possibility that sample 6 was simply contaminated with the positive control plasmid expressing the blasticidin gene (Figure 3.10). Each sample containing blasticidin-resistant cells was screened by amplifying sequences from inside the blasticidin-resistance coding sequence. The blasticidin resistance gene was successfully amplified from samples 4-7 but not sample 8 (Lanes 1-4, Figure 3.10a). In order to rule out the possibility of contamination in these samples by the positive control plasmid these same samples were screened by PCR using primers in the CMV promoter and the blasticidin resistance gene that would only amplify a product if the positive control plasmid were present. A product was only amplified from sample 7, which confirmed that the positive control plasmid is only present in this sample (Lane 4, Figure 3.10). RT-PCR confirmed that blasticidin was expressed in sample 6 and sample 7 cells (Figure 3.10c). Attempts to screen isolated colonies from the panning experiments for gene disruptions failed to yield any high-confidence colonies containing the CCR5 disruption (data not shown).

3.3.4 Southern Blot Analysis of RNP-Stimulated Homologous Recombination in Targeted K562 Cells

To further characterize sample 6, which was shown to have a detectable integration event by PCR, I designed a Southern blot strategy to confirm the integration at the CCR5 target site (Figure 3.11a). This strategy involved the use of a DNA probe designed to anneal to the CCR5 genomic sequence just upstream of the 5' end of the 5' homology arm (probe 1, Figure 3.11a). Genomic DNA extracted from cells of sample 6 was digested with *Afl*II. The genomic DNA from non-targeted cells would result in a 5.5-kb DNA fragment, while successful integration would yield a DNA fragment 8.7 kb in length (Figure 3.11a). The results of the Southern blot were complicated by the fact that the genomic DNA of sample 6 was extracted from a heterogeneous pool of cells collected prior to panning for isolated antibiotic resistant cells. Therefore, when blotting with the probe designed to anneal to the CCR5 gene just upstream of the targeting event, a band corresponding to the expected size of the integrated targeting vector was not detectable (Figure 3.11b). In other words, the proportion of targeted genomic DNA was insufficient for detection among the heterogeneous WT (non-targeted) DNA.

Subsequently, I designed a second probe to anneal to the blasticidin resistance gene (probe 2, Figure 3.11a). This probe would verify the presence of the blasticidin ORF in the genomic DNA. The Southern blot yielded a single band in each lane containing genomic DNA from blasticidin resistant cells (Figure 3.11c). However, I was unable to distinguish the band corresponding to the genomic integration (8.7 kb) from the expected sized band for the linearized vector alone, which, albeit unlikely, may be present to some degree as extrachromosomal DNA (9.6 kb) (Figure 3.11c). A number of

experiments were performed to isolate a single CCR5-targeted cell colony from sample 6 to confirm that individual cells were positive for the gene disruption, but no conclusive evidence was found (data not shown). Because the targeting efficiency using dsDNA as the donor DNA is so inefficient and site-specific double strand cleavage by the group II intron is also thought to be inefficient, we sought other methods of gene targeting that would improve our likelihood of detecting gene targeting events. We found the rAAV-based gene targeting method to be appropriate for these experiments.

3.4 Targeting the CCR5 Gene *In Vivo* by Group II Intron RNP-Stimulated Integration of Recombinant Adeno-Associated Virus

While introducing exogenous plasmid DNA as described above for site-specific homologous recombination has been a historically effective technique for gene targeting, the most innovative method for contemporary gene disruption is the use of recombinant adeno-associated viruses. It has been shown that rAAV-based vectors integrate into homologous chromosomal loci 25-fold more efficiently than comparable plasmid vectors (Yanez & Porter 1998, Topaloglu et al. 2005, Russell & Hirata 1998). With gene targeting efficiencies approaching 70% in some cases of rAAV targeting, we felt that utilizing rAAV would provide a reliable positive control level of baseline targeting upon which we could demonstrate a stimulatory effect attributable to double-strand breaks created by group II intron RNPs at the target site (Hirata et al. 2002, Kohli et al. 2004, Topaloglu et al. 2005, Porteus et al. 2003). It has been shown previously that the introduction of a DNA double-strand break in a DNA target gene can stimulate the frequency of rAAV-mediated gene targeting over 100-fold (Porteus et al. 2003, Miller et al. 2003). Accordingly, we chose to adapt this technique for our purposes of verifying

that group II intron RNPs can introduce double-strand breaks at the genomic level capable of stimulating the integration of rAAV-based gene targeting vectors (Porteus et al. 2003). To do so, we obtained rAAV gene targeting vectors for the CCR5 gene and applied them to HCT116 cells along with the CCR5-targeted group II intron RNP.

3.4.1 Designing the rAAV Targeting Vectors for the CCR5 Gene

Designing a rAAV targeting vector basically involves the replacement of the viral genome of the WT AAV between the *cis*-acting terminal repeats with foreign DNA (Hirata & Russell 2000). Removing the viral genes, specifically the *rep* gene, prevents insertion at the site-specific integration locus of wild-type AAV located on human chromosome 19 and allows for the insertion of new genetic cargo material to be delivered by the virus to the cell. Foreign DNA inserted in place of the viral genome usually consists of a transgene cassette containing a promoter and polyadenylation signal most often totaling less than 4.7 kb in length, limited specifically by the cargo capacity of the encapsidated virus. AAV vectors are directed to insert into gene-specific locations in the genome by DNA sequences with homology to the chromosome. Interestingly, rAAV vectors can integrate site-specifically using four times less homologous sequence than conventional plasmid-based homologous recombination-dependent targeting vectors (Vasileva & Jessberger 2005). This reduces the need for long regions of homologous sequence flanking the selection marker.

To disrupt the CCR5 gene in HCT116 cells, we used two different rAAV targeting vectors. The rAAV construct used in our first rAAV-based gene targeting experiments is described in Kohli et al. (2004), and was obtained from Bert Vogelstein.

The rAAV expression vector, pAAV-CCR5, is designed for expression of the viral genome using the accompanying Stratagene vectors (Figure 3.6b). The original intent for this vector in the Vogelstein lab was to re-create the CCR5 Δ32 deletion identified in certain human populations that have been shown to naturally exhibit resistance to HIV-1 infection (Figure 3.6b). By chance, the 5' homology arm (HA) of pAAV-CCR5 overlaps the target site for the CCR5-targeted group II intron RNP (Figure 3.6b). Specifically, the 5' HA begins 596 bp upstream of our group II intron insertion site and extends 231 bp downstream of it (Figure 3.6b). The 3' HA begins 277 bp downstream of our target site and extends 1184 bp downstream of it (Figure 3.6b). Replacing the viral genome are the homology arms that flank a gene conferring resistance to neomycin (G418) linked in tandem to a gene conferring zeomycin resistance. The *neo*^{R} gene is driven by the phosphogycerate kinase (PGK) promoter and also contains a polyadenylation signal (pA) (Figure 3.6b). Use of a selection marker cassette containing an exogenous promoter offered a much-improved opportunity for expression of the selection marker than when depending upon the endogenous promoter from the CCR5 gene. The zeomycin-resistance gene derived from pZeoSV contains an EM7 promoter and is used for selection in bacterial cells. Also present in the targeting vector upstream of the *neo*^{R} gene and downstream of the zeomycin resistance gene are *loxP* sequences for removal of the antibiotic resistance genes after genomic integration (Kohli et al. 2004). Because this rAAV gene targeting vector is designed to integrate in the region of the CCR5 gene containing our target site, we had reason to believe that a double-strand break introduced

by the CCR5-targeted group II intron RNP could stimulate incorporation of the vector at the target site.

Moehle et al. (2006) demonstrated that conventional plasmid-based targeting vectors containing 750 bp of homologous sequence directly flanking an artificially induced DNA double-strand break is sufficient to deliver up to 7.7 kb of exogenous DNA into the genome. With this in mind, we designed a second rAAV vector that fit these parameters. Based on the results presented in Moehle et al. (2006), we designed the vector to contain 750 bp of homologous sequence on either side of the CCR5-targeted group II intron cleavage site (Figure 3.6c). Carlo Rago from the Vogelstein lab constructed the second rAAV vector (pAAV-CCR5t) using previously described methods (Kohli et al. 2004, Rago et al. 2007). The homology arms flank a selection marker cassette similar but not identical to the cassette previously described for pAAV-CCR5 (Figure 3.6b). In the newly designed rAAV vector, a SV40 promoter rather than a PGK promoter drives expression of the *neo*^{R} gene (Figure 3.6c). A multi-cloning site (MCS) was also added upstream of the SV40 promoter and downstream of the *loxP* sequence. Although these differences are significant at the sequence level, the selection cassette in pAAV-CCR5t is expected to perform similarly to that for the pAAV-CCR5 vector.

3.4.2 RNP-Stimulated Integration of a Recombinant AAV Vector

The method of targeting the CCR5 gene by group II intron RNP-stimulated homologous recombination in conjunction with the rAAV vector requires adaptation of the conventional rAAV gene targeting approach (Rago et al. 2007). Because these experiments were performed using HCT116 cells, a commonly used cell line for rAAV

experiments, we were able to modify existing protocols established in the Vogelstein lab for targeting these cells with rAAV alone (Rago et al. 2007). The essential difference in the targeting experiment is the introduction of the group II intron RNP. Addition of the RNP to the targeting process required modifications to the protocol that allowed for delivery of the RNP subsequent to transduction with the rAAV virus. The most established method in our lab for delivering the RNP to the cells is by electroporation, as previously described. For HCT116, there was no established electroporation protocol for the delivery of RNPs so I tested electroporation conditions previously utilized for HEK293 cells and K562 cells (Table 3.1).

Briefly, HCT116 cells were grown in T-25 cell culture flasks to approximately 60-80% confluency in McCoy's 5A cell culture medium. All the samples were treated identically leading up to the experiment. Once the cells reached the desired confluency, the medium was removed and the cells were washed with PBS. Four different types of samples were included in a typical experiment. Sample-type 1 received no virus and no RNP. Sample-type 2 received the virus but was not treated with RNP, nor was this sample electroporated. Sample-type 3 received virus and received an electroporation pulse without any RNP. Sample-type 4 was treated with virus and electroporated with the CCR5-targeted RNP. Variations on sample 4 usually included differing amounts of RNP and/or different electroporation conditions.

As described in Rago et al. (2007), the rAAV virus was added drop-wise to the surface of the cells in the T-25 flask. Immediately following the addition of the virus, the cells were covered with 2.5 mL of Opti-MEM (Invitrogen). The cells were incubated

with the virus for 2 hours at 37ºC. After incubation with the virus, the cells were treated according to prescribed conditions with or without the group II intron RNP and with or without electroporation. Accordingly, sample 1 usually served as the negative control receiving no virus, no RNP, and no electroporation. For this sample, cells were simply washed with PBS prior to adding 4 mL of complete growth medium and incubation at 37ºC. Sample 2, the positive control sample for viral targeting, received 4 mL of complete growth medium after viral transduction and incubation at 37ºC. Samples that were designated to receive RNP were rinsed with PBS and trypsinized to release them into suspension. The cells were harvested by centrifugation and resuspended in complete medium to neutralize the trypsin. Cells from samples 3 and 4 were washed 3 times with FBS (-) medium. The cells were counted and subsequently diluted to 1.25 x 10^6 cells/mL. Next, aliquots of 800 μL of the cell suspension were transferred to Eppendorf tubes corresponding to each sample 3 and 4. Cells for sample 3 were transferred to the electroporation cuvette to be electroporated according to established conditions (Table 3.1). A select amount of RNP, usually 10-50 μg, was added to sample 4 and the cells were electroporated, as described above. Following electroporation, the cells were immediately transferred to a T-25 flask containing 2.5 mL of FBS (-) medium. All the samples were incubated at 37ºC for 72 hours. After 72 hours, each flask of cells was trypsinized. The cells were collected by centrifugation and resuspended in fresh complete McCoy's 5A medium containing 400 μg/mL Geneticin for antibiotic selection of resistant cells. The cells were diluted to 10^5 cells/mL and 100 μL (10^4 cells total) of each sample was distributed to each well of 96-well plates. The plates were wrapped in

plastic wrap and incubated for 14 days at 37ºC. After 14 days, wells containing grown cells were trypsinized and the antibiotic-resistant cells were transferred to a second 96-well plate such that each well contained cells resistant to antibiotic selection. Once the cell growth in the new plates reached approximately 80% confluency, genomic DNA was extracted using a 96-well Blood Kit (QIAgen). DNA concentration was determined using a nanodrop and the eluted DNA was used as the template for PCR screening for gene targeting events.

3.4.3 PCR Analysis of RNP-Stimulated Targeting by Recombinant AAV at the CCR5 Target Site

The PCR-based strategy for identifying the cell clones that contain recombinant alleles involved the amplification of a DNA fragment using a primer that annealed within the Neo^R gene and a second primer that anneals in the CCR5 gene outside the homology region of the rAAV vector (Figure 3.12). The resulting amplicon of known size spans the entire HA and will only be amplified when the marker is integrated at the target site. PCRs were carried out in 20 μL reaction volumes using modified parameters from Rago et al. (2007). The PCR products were separated by gel electrophoresis and lanes were examined for the presence or absence of a PCR product of the correct size. Because the homology arms used for targeting were ~900 bp, the complete amplicon, which consisted of the homology arm and a portion of the *neo*^{R} expression cassette totaled 1.6 kb (Figure 3.12b). Presence of the correct sized PCR product indicated that the specific targeting event had occurred at the CCR5 target site. PCR products were sequenced for confirmation.

The targeting experiment in which pAAV-CCR5 was the plasmid source for production of the rAAV vector was expected to be sub-optimal for observing a stimulatory effect of the CCR5-targeted intron on integration, because the 5' homology arm overlapped the double-strand cleavage site rather than flanking it. Despite these circumstances, I was able to detect targeting events with a high degree of site-specific integration among all samples that received the virus. Figure 3.12c summarizes the results of the targeting for each of the 4 samples in the pAAV-CCR5 targeting experiment. A typical PCR screening panel is depicted in Figure 3.12b. As can be seen in Figure 3.12c, samples receiving the electroporation pulse with and without the RNP exhibited a higher targeting efficiency than those that did not receive the pulse. It has been found previously that the ratio of targeted to random integrations using double-stranded DNA targeting vectors can be increased based on the voltage and capacitance used during electroporation (Mohn & Koller 1995). While electroporation alone did increase the targeting efficiency 1.7-fold above that observed for the non-electroporated sample (from 36% to 64%), cells that received the CCR5-targeted RNP (50 μg) showed an even higher targeting efficiency (81.25%), 2.2-fold above that observed for the non-electroporated sample (Figure 3.12c). The additional increase in targeting efficiency associated with the inclusion of the group II intron RNP in the targeting experiment suggests that the RNP may have an enhancing effect on rAAV integration. We would expect that double-strand cleavage at the target site by the group II intron RNP would stimulate the integration of the rAAV, however we had no way of predicting the extent of the stimulation. In this case, it might be argued that the group II intron RNP provides an

increase of 120% over the targeting efficiency for cells that received the virus without electroporation or the RNP. Further, we observed an increased targeting efficiency 27% above that for the electroporated cells that received the virus but no RNP (Figure 3.12c). Of course, these results can be further scrutinized and it is recommended that these experiments be repeated. However, since this particular rAAV vector is believed to be sub-optimal, it was not in our interest to pursue these experiments further.

In another targeting experiment we used the customized rAAV vector, produced from pAAV-CCR5t, which was designed in the optimal configuration for integration at the target site compatible with the CCR5-targeted group II intron such that the homology arms directly flank the double-strand cleavage site (Moehle et al 2007) (Figure 3.6c). Figure 3.12d summarizes the results of this experiment in which two different amounts of RNP were applied in conjunction with two different electroporation conditions for the delivery of the RNP. Once again the samples receiving the electroporation pulse had fewer colonies than the samples that did not receive the pulse. Also, cells electroporated with the higher amounts of RNP (18 μg) showed increased targeting efficiencies (100% and 93.3%) over those that received no RNP or 5 mg of RNP (83.3% and 75%, respectively) (Figure 3.12). In this experiment, two different amounts of RNP were delivered using two different electroporation conditions. Figure 3.12d shows that the samples receiving a lesser amount RNP had lower targeting efficiencies. Further, the sample that was treated with the higher amount of RNP and electroporation condition 1 exhibited a higher targeting efficiency than the sample that received the same amount of RNP and electroporation condition 2. This suggests that electroporation condition 1

more efficiently delivered the RNP than electroporation condition 2. As shown in Figure 3.12c, both samples containing the higher amount of RNP, regardless of the electroporation condition, showed a higher targeting efficiency than the sample that received no CCR5-targeted RNP. Similar to the previous experiment, the targeting results suggest that electroporation of the HCT116 cells with the RNP has a stimulatory effect on the integration of the rAAV, presumably due to double-strand cleavage by the RNP at the target site.

3.5 Discussion

The need for new gene-targeting technologies is clear. The current gene targeting tool kit is rife with weak and troubled technologies. Even the most recently developed and commonly used methodologies, such as siRNAs and zinc-finger nucleases, have significant drawbacks that limit their versatility. For siRNAs, the need to constantly replenish the RNAs requires plasmid DNA to be transfected into the cells, which can result in uncontrollable integration, presenting potentially dangerous repercussions (Ledwith et al. 2000, Robertson 1994). Such risks of random integrations are the reason some gene therapy programs have been discontinued (Hacein-Bey-Abina et al. 2003). As for zinc-finger nucleases, their value has yet to be determined, largely because the need for protein engineering has limited the production of a broad selection of targeted nucleases. The number of genes that have been targeted using ZFNs is disappointing given the amount of time that has passed since the technology was shown to be effective for gene targeting. Until the engineering of zinc-finger nucleases becomes more simplified and more productive, the uses of ZFNs will be limited.

The absence of a perfect gene-targeting method highlights the necessity for developing new technologies like that proposed here for group II introns. Advancing group II introns toward gene targeting will provide a new tool with a broad range of biochemical activities that can be exploited for a variety of genetic engineering applications. The ability to deliver genes, disrupt genes, and act as site-specific endonucleases makes group II introns a powerful alternative to existing technologies. The fact that this single particle can carry out so many different processes makes it a potentially invaluable option with seemingly infinite applications.

In this chapter, I provide data to show that group II introns should be considered a legitimate prospect for future gene targeting in eukaryotic cells. Obviously, a significant amount of development remains to be done, but based on the results presented here, group II introns have potential to be used in at least two different ways to disrupt target genes. First, I present evidence that group II introns have the ability to integrate into the genome of cultured mammalian cells. Using electroporation, group II introns were delivered to K562 cells. From these cells, PCR amplified a fragment corresponding to the 3' integration junction of the CCR5-targeted intron at the CCR5 target site. Integration at a single-copy endogenous gene has not been described previously. Repeating these experiments and determining the optimal conditions for single copy genes would be a huge leap forward in our gene targeting capabilities. Second, I used two different types of gene targeting vectors designed to disrupt the CCR5 gene by homologous recombination to show that a group II intron could stimulate their integration and enhance the targetability of these vectors. A linearized plasmid DNA designed to

integrate into the CCR5 gene was delivered with and without the CCR5-targeted RNP. Only those cells that received both the targeting vector and the RNP successfully integrated the donor DNA. PCR confirmed the integration of the vector at the target site. The apparent RNP dependence for the integration of the donor DNA was a promising result and provided a novel protocol for using the RNP in eukaryotic cells. To reinforce the results demonstrated using the double-stranded DNA targeting vector, I obtained two different rAAV vectors designed to disrupt the CCR5 gene. The rAAV vector utilizes the single-stranded genome of the virus to integrate into the genome in a site-specific manner based on sequences that have homology to the target site. The two rAAV constructs used in these studies differed such that they were designed to integrate at different sites in the CCR5 gene. The sites were less than 300 bp apart and the homology arms of the vectors either overlapped or flanked the cleavage site for the group II intron RNP. For both vectors, it was determined that the group II intron RNP could stimulate the integration of the recombinant viral genome. We conclude that the stimulation of homologous recombination is likely the result of double-strand breaks created by the RNP. These results are consistent with other experiments that show that double-strand breaks can stimulate the integration of rAAV, provided they occur in proximity to the rAAV target sequence (Porteus et al. 2003, Miller et al. 2003).

While I have high confidence that group II introns stimulated the integration of both the double-stranded DNA and the rAAV gene targeting vectors, the activity of the RNP appears to be inefficient when compared to similar experiments performed using other types of site-specific endonucleases (Moehle et al. 2007, Porteus et al. 2003).

Whereas ZFNs were shown to stimulate the integration of double-stranded DNA by as much as 250-fold and rAAV by 100- to 300-fold, group II introns were only able to do so by approximately 0.25- to 2-fold (Moehle et al. 2007, Porteus et al. 2003). The explanation for the inefficient activity of group II introns in eukaryotic cells has been discussed in previous chapters. Primarily, the intracellular concentration of Mg^{2+} limits the activity of the RNP. Based on *in vitro* experiments, group II introns function maximally at 10 mM Mg^{2+} while mammalian intracellular *free* Mg^{2+} levels are usually between 0.5 mM and 1.2 mM (Romani & Scarpa 2000, Romani & Maguire 2002). Further, plasmid targeting in zebrafish and *Drosophila* embryos and *Xenopus* oocytes was also shown to require intracellular Mg^{2+} concentrations that approached 10 mM for optimal intron integration (Mastroianni et al. 2008). Unfortunately, increasing the intracellular Mg^{2+} levels in mammalian cells is difficult because mammalian intracellular Mg^{2+} concentrations are highly regulated and large increases in Mg^{2+} concentrations would likely result in significant cell death (Romani & Scarpa 2000, Patel et al. 1994, Hartwig 2001). Alternatives to addressing the Mg^{2+} limitations might include engineering a group II intron or identifying a naturally occurring retargetable group II intron that functions at low Mg^{2+}.

It is also likely that the inefficient targeting observed in the experiments presented here is due to inefficient delivery of the RNP by electroporation. Developing an assay that tests the effective delivery of the RNP or repeating the experiments presented here using a broad range of electroporation conditions could overcome this problem. Other possibilities for improving the delivery of the RNP include *in vivo* expression of the

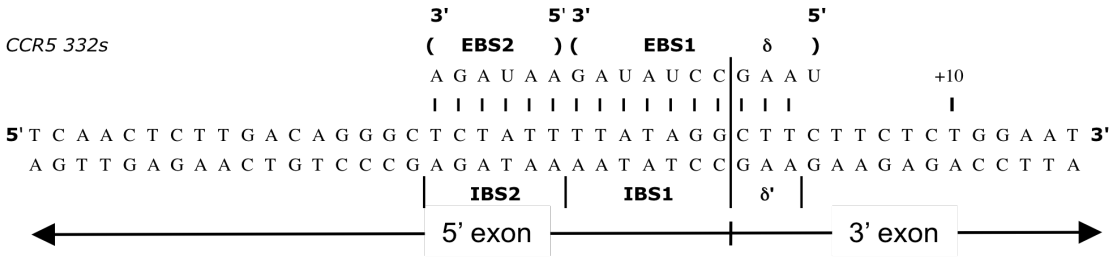
group II intron. Presumably, producing high levels of the group II intron in the cell through expression will increase the probability of targeting. However, as previously mentioned, when expressing the intron *in vivo*, one will have to address the issue of sub-optimal Mg^{2+} for both the forward splicing and reverse splicing reactions. Although preliminary data in our lab suggests that LtrA-dependent forward splicing occurs in mammalian cells when both the protein and the intron RNA are expressed from a plasmid, the splicing efficiencies were found to be low. Again, this should refocus optimization efforts on addressing the problem of low intracellular Mg^{2+} concentrations.

Another limiting factor for group II intron gene targeting in mammalian cells that may reduce the efficiency of gene targeting is nuclear localization. While preliminary data performed by others and the data presented here suggest that electroporated group II introns are able to localize to the nucleus of mammalian cells to find their targets, closer examination will likely reveal that the efficiency of nuclear localization is low. Electroporation is effective for delivering nucleic acids and RNPs to cells, however reaching the nucleus is much less efficient than reaching the cytoplasm and often requires active transport (Favard et al. 2007, Lechardeur & Lukacs 2002, Golzio et al. 2002). Further, despite the fact that others in our lab have shown through expression analysis that attaching a nuclear localization signal to LtrA drives the protein component of the RNP to the nucleus, it is unclear whether the protein is delivered with or without the intron RNA. Testing RNP delivery using equipment such as the Nucleofector (Amaxa), which is designed to deliver nucleic acids specifically to the nucleus with higher

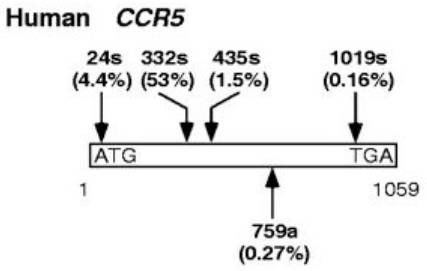
efficiencies than standard electroporation devices, may assist in addressing the issue of nuclear localization (Martinet et al. 2003).

The experiments described here address specific limitations of experiments previously performed in our lab. Some of the drawbacks of previous experiments performed in our lab include: 1) the lack of a positive control for experimental comparisons, 2) the inefficiencies associated with the group II intron targeting activity in eukaryotic cells, especially regarding single-copy target sites, and 3) the inability to enrich for and isolate living cells containing group II intron-based targeting events. I set out to design experiments that would address each of these issues. Use of rAAV as a gene-targeting vector with measurable integration efficiency offered a positive control level of targeting upon which the enhancement effects of the group II intron can be quantified. In addition, the goal of enriching and isolating cells containing gene-targeting events was accomplished using both the integration of the linearized double-stranded donor DNA and the integration of the rAAV targeting vectors, which contained selection markers. Of the two types of targeting vectors, rAAV proved to be a more useful tool for determining the effects of group II introns on integration because rAAV integrates with a quantifiable frequency. Increases in targeting efficiency may be attributed to the group II intron. Thus, rAAV is more amenable to the necessary optimization experiments that will need to be performed in the future to learn more about the targeting behaviors of the group II intron *in vivo*. It is my great hope that the advancements toward genome targeting described here will be influential in designing future experiments to improve the targeting efficiency of group II introns in eukaryotic cells.

(a)



(b)



(c)

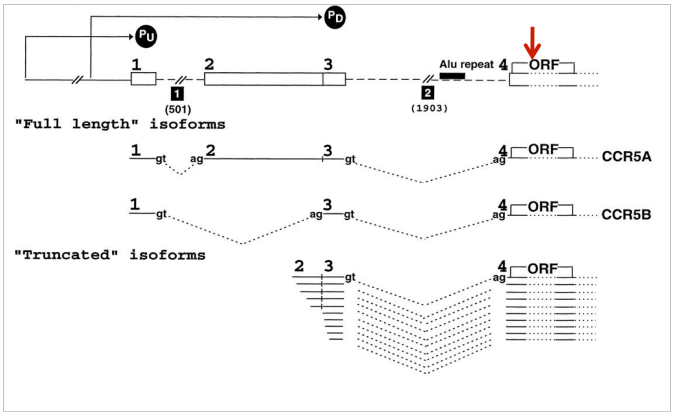


Figure 3.1 Target Site for the CCR5-Targeted Group II Intron

(a) Alignment of the intron EBS1, EBS2 and δ sequences with the DNA target site sequences (IBS1, IBS2 and δ') in the human genomic sequence for the CCR5 gene. The CCR5-targeted intron was among the first customized introns to be designed using data gathered from a library of randomized DNA target sites (IBS and δ') (Guo et al., 2001). (b) Among several introns designed to insert into the CCR5 gene, CCR(332s) had the highest mobility (53%). The intron inserts into the CCR5 gene 332 nucleotides downstream of the start codon. Adapted from Guo et al. (2001). (c) The insertion site (red arrow) is located in exon 4 of the CCR5 gene (arrow). Adapted from Mummidi et al. (1997).

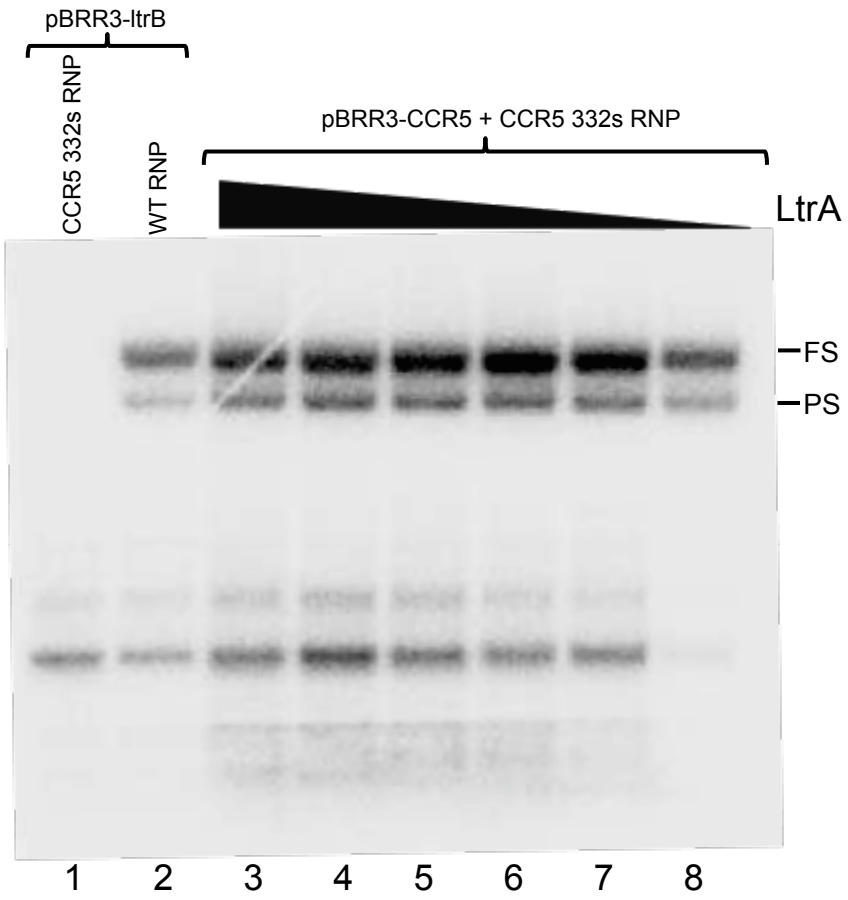


Figure 3.2 TPRT Results to Determine the Optimal Ratio of LtrA Protein to Intron RNA for the Preparation of CCR5 332s Group II Intron RNPs

While it has been shown that the binding ratio of purified LtrA to purified lariat intron RNA is 2:1 (Saldanha et al., 1999), the preparation of RNPs according to methods described here required empirical determination of this ratio due to the presence of extraneous RNAs included in the reconstitution reaction. The RNP particles were prepared by combining purified LtrA protein and excised lariat intron RNA obtained by *in vitro* self-splicing of the CCR5-targeted Ll.LtrB intron RNA (pACD2-CCR5(332s)/*Nhe*I *in vitro* transcript). The spliced RNA included in the RNP reconstitution contains a mixture of RNAs, 60-70% of which is intron lariat, while the remaining 30-40% consists of precursor RNA and splicing products including ligated exons and linear intron RNA. Therefore, when determining the RNA concentration by $O.D._{260}$, it is difficult to specify the exact amount of protein to include in the RNP reconstitution. The TPRT results shown here reflect the *in vitro* targeting activity for RNPs reconstituted with increasing amounts of LtrA (20 nM, lane 8, 40 nM, lane 7, 60 nM, lane 6, 80 nM, lane 5, 120 nM, lane 4, and 160 nM, lane 3) and a constant concentration (20 nM) of spliced intron RNA (83 μg). The optimal targeting activity Lane 6 corresponded to a ratio of 60 nM protein to 20 nM RNA which is a 3:1 ratio.

(a) (b)

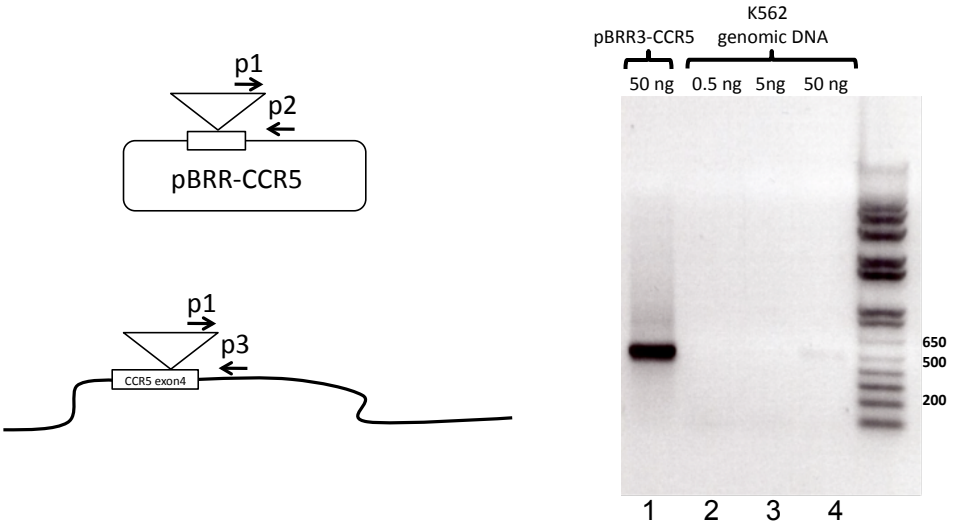


Figure 3.3 PCR to Detect Intron Integration into CCR5 Genomic DNA Target Site in Cold TPRT Reactions

Targeting reactions were carried out by incubating 1 μg of the unlabeled plasmid DNA, pBRR3-CCR5, (Lane 1) or specified amounts (0.5 ng, Lane 2, 5.0 ng, Lane 3, 50 ng, Lane 4) of genomic DNA with 1 μg of reconstituted CCR5-targeted RNP particles in 20 μL of TPRT buffer [10 mM KCl, 10 mM $MgCl_2$, and 50 mM Tris-HCl (pH 7.5)] containing dATP, dGTP, dTTP and dCTP (0.2 mM each) (Invitrogen). The reactions were initiated by addition of the RNP particles, and the mixtures were incubated for 30 min at 37°C. Reactions were stopped by phenol-CIA extraction. The DNA was ethanol precipitated from the supernatant. The DNA pellet was resuspended in 10-20 μL of TE [10 mM Tris-Cl, pH 7.5, 1 mM EDTA] and used as the DNA template for PCR. (a) For plasmid targeting, the 3'-integration junction was amplified by PCR using the intron-specific primer, ltrB 816s (p1), and a plasmid-specific primer located downstream of the target site, Seq pBRR MCS+RR (p2). The expected size of the amplicon was 586 bp. For genomic DNA targeting, the 3'-integration junction was amplified by PCR with the intron-specific primer, ltrB 816s (p1), and a CCR5-specific primer located downstream of the target site, CCR5 +367-+340 (p3). The expected size of the amplicon was ~492 bp. (b) Successful amplification across the 3' junction suggests that the intron has correctly inserted at its target site and the intron RNA is reverse transcribed at least up to the annealing point for the PCR primer. The 586-bp PCR product corresponding to the 3' junction was readily amplified from the plasmid targeting experiment that was performed parallel to the genome targeting experiment. The 492-bp amplicon corresponding to the 3' junction for genomic DNA targeting verified that the CCR5 RNP could find its target site even within a more complex genetic environment. Based on the amounts of genomic DNA used in the cold TPRT reactions (50 ng, 5.0 ng and 0.5 ng), *in vitro* targeting of the CCR5 gene could be detected on an agarose gel in as few as 1,510 genomes.

Table 3.1 Electroporation Conditions for RNP Delivery.

Cell Type	Voltage (V)	Capacitance (uF)	Cell Number	Cell Concentration	Buffer
HEK293	230	250	$2x10^6$	$5.0x10^6$	PBS
K562	320	800	$1x10^6$	$5.0x10^6$ or $1.25x10^6$	IMDM
HCT116 Condition A	230	250	$1x10^6$	$1.25x10^6$	McCoy's 5A
HCT116 Condition B	320	800	$1x10^6$	$1.25x10^6$	McCoy's 5A

The electroporation conditions used to deliver group II introns to three different cell lines are described in this table. The conditions for HEK293 cells were empirically determined in our lab. The conditions for K562 cells were based on Bio-rad electroprotocols for delivery of DNA. HCT116 cells were electroporated using the protocols for both HEK293 and K562 cells.

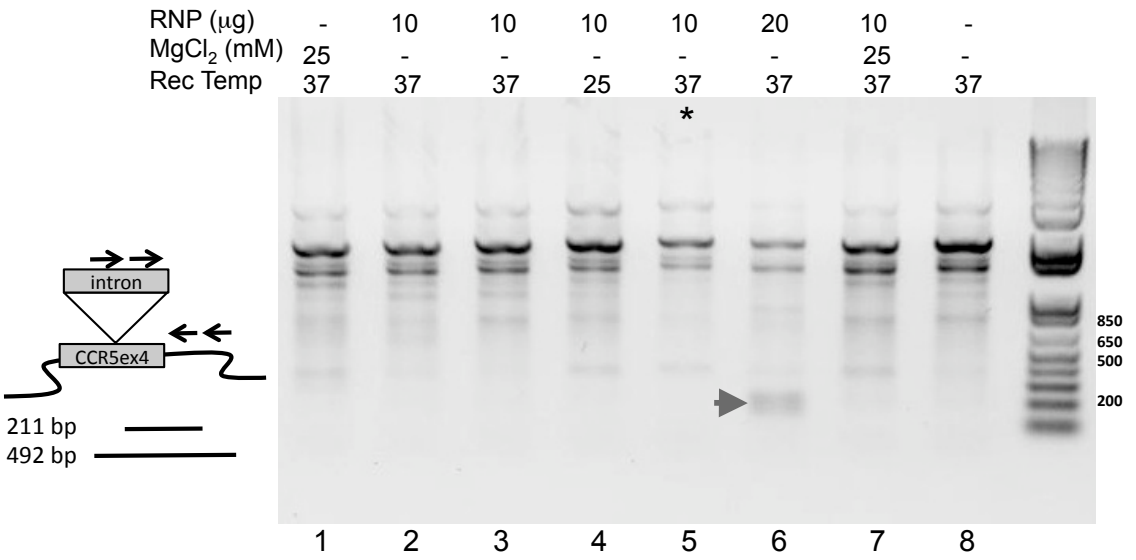


Figure 3.4 Direct Integration of a Group II Intron at the CCR5 Target Site

The CCR5-targeted RNP was delivered to K562 cells to disrupt the CCR5 gene by direct integration. K562 cells were grown in IMDM (Invitrogen) with 10% FBS and penicillin-streptomycin (Invitrogen) to 1×10^6 cells/mL. Cells were harvested by centrifugation and washed 3 times with FBS (-) IMDM. The cell pellet was resuspended in FBS (-) IMDM to a density of 6.25×10^6 cells/mL or 6.25×10^5 cells/mL. 5×10^6 cells (800 μL) or 5×10^5 cells (800 μL) were transferred to 1.5 mL-microcentrifuge tubes and combined with specified amounts (10 μg or 20 μg) of CCR5-targeted RNPs at room temperature. 2 M $MgCl_2$ was added to some samples to a final concentration of 25 mM. The cells were electroporated using a Gene Pulser II electroporation system (Bio-Rad). After electroporation, the cells were incubated at either room temperature or 37°C for 15 min and transferred to a T-25 flask containing 10 mL of IMDM complete medium. Genomic DNA was extracted using the GenElute Mammalian Genomic DNA Miniprep Kit (Sigma). Intron integrations were amplified by PCR to detect the 3'-integration junction. The primary PCR used the intron-specific primer, ltrB 816s, and a CCR5-specific primer, CCR5 +367-+340. Primary PCR products were purified using the QIAquick PCR Purification Kit (Qiagen). The purified PCR products served as the template for nested PCR using primers, ltrB 890s and CCR5 +156-+140. The results are shown in the 1% agarose gel above in which lanes 1-8 correspond to the following samples: Lane 1, No RNP + 25 mM $MgCl_2$, $5x10^6$ cells, 37°C recovery; Lane 2, 10 μg RNP, $5x10^6$ cells, 37°C recovery, No electroporation pulse; Lane 3, 10 μg RNP, $5x10^6$ cells, 37°C recovery; Lane 4, 10 μg RNP, $5x10^6$ cells, 25°C recovery; Lane 5, 10 μg RNP, $5x10^5$ cells, 37°C recovery; Lane 6, 20 μg RNP, $5x10^6$ cells, 37°C recovery; Lane 7, 10 μg RNP + 25 mM $MgCl_2$, $5x10^6$ cells, 37°C recovery; and Lane 8, No RNP, $5x10^6$ cells, 37°C recovery. Intron integration was detected in a single sample (Lane 6) in which 20 μg of RNP was delivered to the cells.

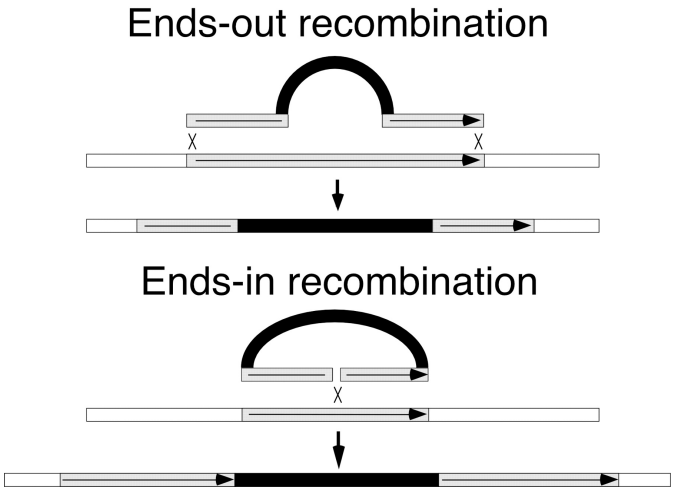
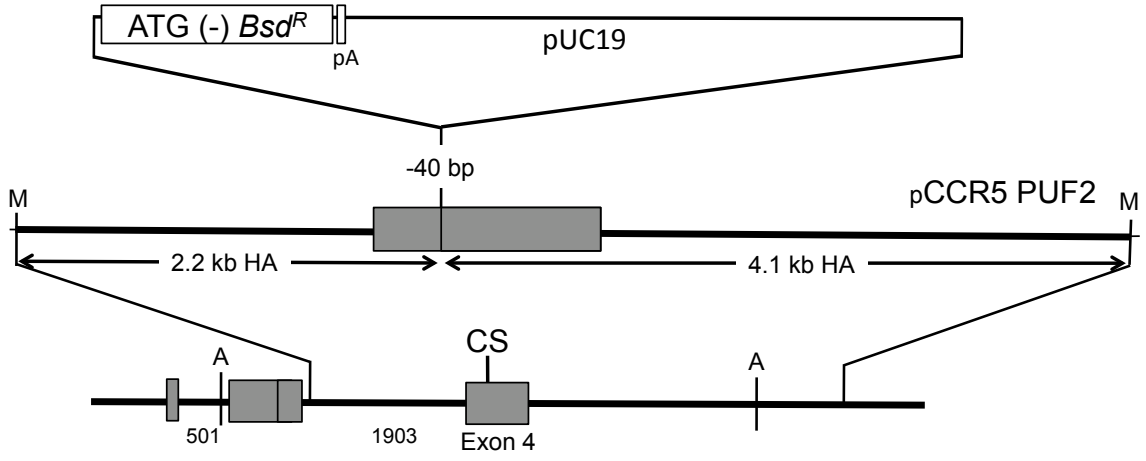


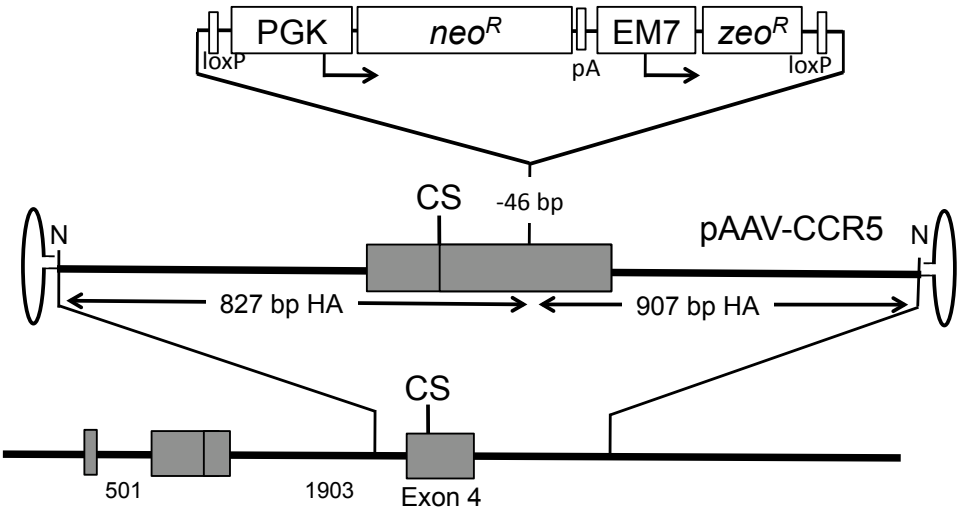
Figure 3.5 Configurations for Plasmid-Based Gene Targeting Vectors

Donor DNA molecules fall into one of two categories. Either they are classified as "ends-out" or "ends-in" recombination constructs. Ends-out constructs are also known as replacement constructs because they can replace genomic sequences while inserting new sequence at the site of integration. The experiments performed here employ the ends-out form of gene targeting vector. Adapted from Gong & Golic (2003).

(a)


ATG (-) Bsd^R
pA
pUC19
-40 bp
M
pCCR5 PUF2
M
2.2 kb HA
4.1 kb HA
CS
A
A
501
1903
Exon 4

(b)


PGK
neo^R
EM7
zeo^R
loxP
pA
loxP
CS
-46 bp
N
pAAV-CCR5
N
827 bp HA
907 bp HA
CS
501
1903
Exon 4

(c)

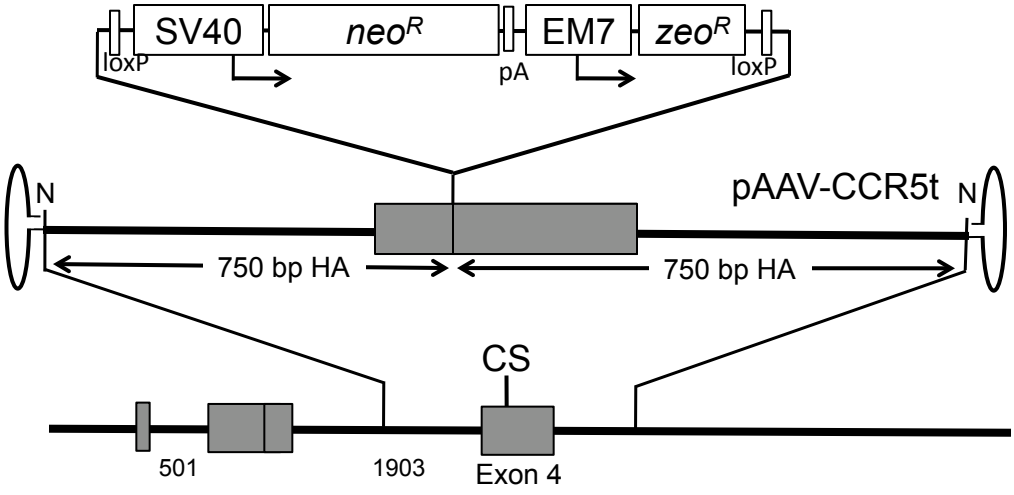

SV40
neo^R
EM7
zeo^R
loxP
pA
loxP
N
pAAV-CCR5t
N
750 bp HA
750 bp HA
CS
501
1903
Exon 4

Figure 3.6 Donor DNA Constructs for Targeting CCR5

(a) The pick-up fragment (PUF) donor DNA, pCCR5-PUF2, was designed to incorporate at the CCR5 gene by homologous recombination (HR) in K562 cells. The design of the PUF construct is based on standard HR DNA constructs used in classical gene targeting experiments in which regions of sequence homology flanking the desired insertion site are located on either side of a selection marker (Doetschman, 1988, Bronson, 1994). Two regions of homology were amplified by PCR from the CCR5 gene; the first beginning 25 bp 5' of the target site and extending 2.2 kb upstream, including a portion of intron 2 and exon 4, and a second fragment beginning 15 bp 3' of the target site and extending 4.1 kb downstream, consisting of a portion of exon 4 and intron 3. The 40 bp directly flanking the target site were omitted from the construct in order to prevent the RNP from targeting the donor DNA molecule prior to or following insertion. An ATG (-) blasticidin resistance gene was cloned in frame with the upstream coding sequence of exon 4 from the CCR5 gene. Therefore, expression of the blasticidin-resistance gene would necessarily be driven by endogenous transcription from the CCR5 promoter upon integration. Successful targeting events could then be selected based on blasticidin resistance. The donor DNA was constructed in a pUC19 plasmid backbone such that linearization (*Mlu*I digestion) of the vector was necessary for the targeting vector to take on an "ends-out" arrangement when aligned along the targeted chromosome. Upon integration, the blasticidin selection cassette and the pUC19 vector backbone replace the 40-bp sequence corresponding to the group II intron recognition site upon integration.

(b) To disrupt the CCR5 gene in HCT116 cells, we used two different rAAV targeting vectors. pAAV-CCR5, obtained from Bert Vogelstein, was designed to re-create the well-documented human CCR5 Δ32 deletion which provides resistance to HIV infection. The 5' homology arm (HA) of pAAV-CCR5 overlaps the target site for the CCR5-targeted group II intron RNP. Specifically, the 5' HA begins 596 bp upstream of the group II intron insertion site and extends 231 bp downstream of it. The 3' HA begins 277 bp downstream of the intron target site and extends 1184 bp downstream of the target site. The homology arms flank a gene conferring resistance to neomycin (G418) linked in tandem to a gene conferring zeomycin resistance. The neo^R gene is driven by the phosphoglycerate kinase (PGK) promoter and also contains a polyadenylation signal (pA). The zeomycin-resistance gene derived from pZeoSV contains an EM7 promoter and is used for selection in bacterial cells. Also present in the targeting vector upstream of the neo^R gene and downstream of the zeomycin-resistance gene are *loxP* sequences for removal of the antibiotic resistance genes after genomic integration (Kohli, 2004).

(c) We designed a second rAAV vector to contain 750 bp of homologous sequence on either side of the CCR5-targeted group II intron cleavage site. The homology arms flank a selection marker cassette similar to the cassette previously described for pAAV-CCR5 except a SV40 promoter rather than a PGK promoter drives expression of the Neo^R gene. The selection cassette in pAAV-CCR5t is expected to perform similarly to that for the pAAV-CCR5 vector.

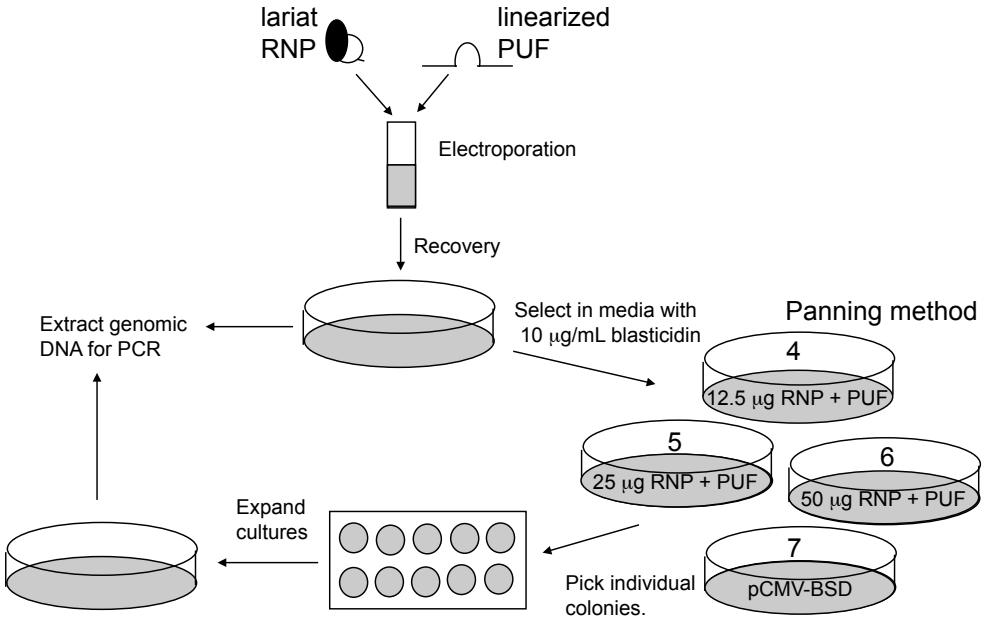


Figure 3.7 Schematic Diagram of the Gene Targeting Method Using a CCR5-Targeted RNP and a Linearized Plasmid Donor DNA

RNPs were delivered along with the linearized pCCR5-PUF2 donor DNA to K562 cells by electroporation using conditions described in Table 3.1 to test for RNP-stimulated integration of the donor DNA. Both the donor DNA and the RNP were resuspended in HKM buffer [10 mM KCl, 10 mM $MgCl_2$, and 40 mM HEPES, pH 8.0]. Cells were suspended at 5 x 10^6 cells/mL in 800 μL of IMDM medium (Invitrogen). A constant concentration of donor DNA (approximately 10 μg) was tested with varying concentrations of CCR5-targeted RNP (12.5 μg, 25 μg and 50 μg). The RNP and DNA were added separately to the cell suspension prior to electroporation. After electroporation, cells were transferred to a T-75 culture flask containing 50 mL of growth medium without FBS and incubated at 37ºC for 1 h. After 1 h, the medium was exchanged, and the cells were incubated overnight in complete medium without selection. After 48 hours of recovery, the selection process for gene targeting events was initiated by adding blasticidin (5 μg/mL) to the medium. Five days post-electroporation, the concentration of blasticidin was increased to 10 μg/mL. Cells resistant to blasticidin were tracked by counting live cells in each of the samples. On day 29, cells were collected and either transferred to plates treated with WSC and conA to pan for live cells or they were used for genomic DNA extractions. Cells used for the panning method were grown up into large enough clusters to pick and transfer to 96-well plates. Genomic DNA extracted from the heterogenous pool of cells was used for PCR screening for PUF integrations. Cells that grew in the 96-well plates were also used for genomic extractions to be screened by PCR for PUF integrations.

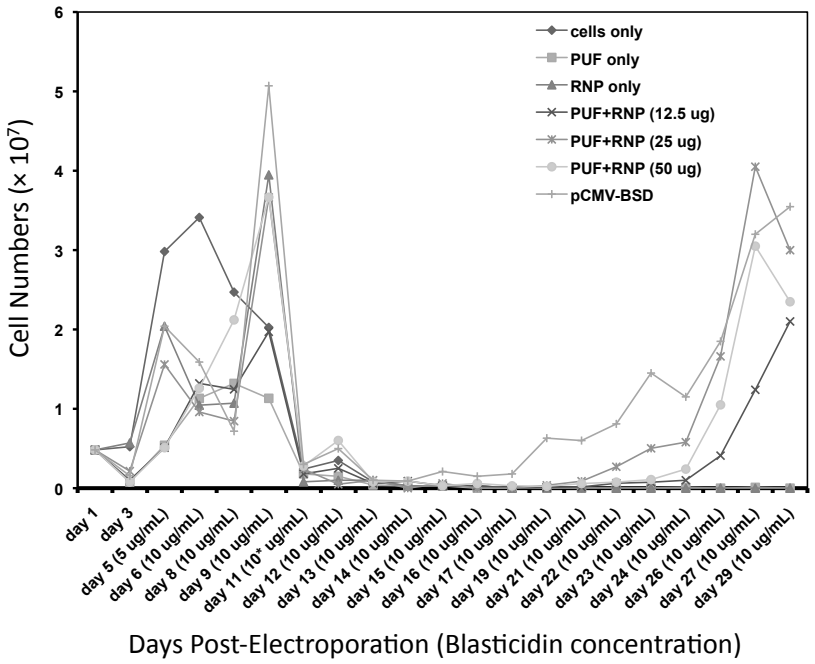
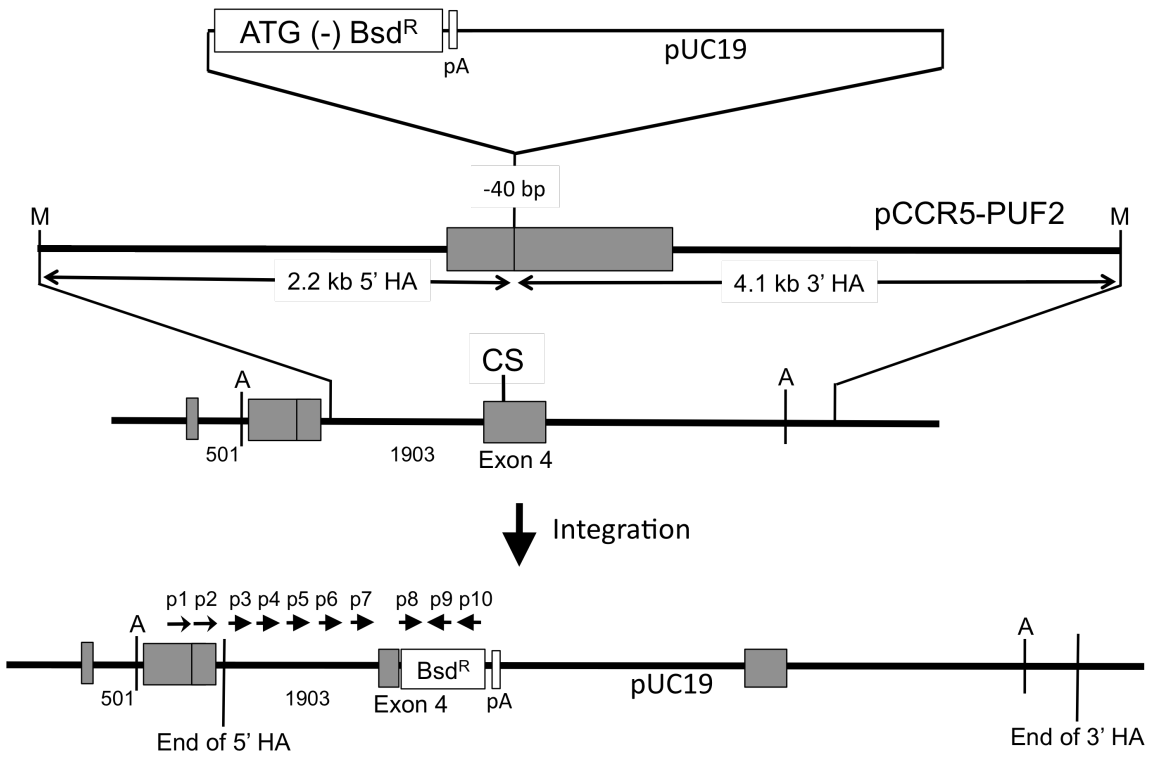


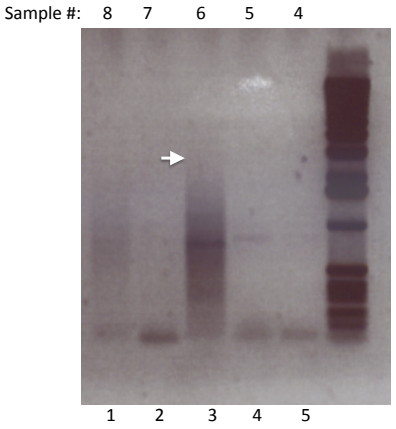
Figure 3.8. Selection of Blasticidin-Resistant Cells Following CCR5 Gene Targeting Using a Plasmid-Based PUF

Following electroporation with the CCR5-targeted RNP and plasmid-based donor DNA, cells resistant to blasticidin were tracked by counting live cells in each of the samples. 48 hours after electroporation, selection of blasticidin resistant cells was initiated by adding 5 μg/mL blasticidin to the medium. Five days post-electroporation, the concentration of blasticidin was increased to 10 μg/mL. By day 11, the cells were experiencing significant cell death in all samples. From day 13 until day 21 post-electroporation, the number of live cells in all samples except the positive control (sample 7) was greatly reduced. The positive control sample had been transfected with a plasmid expressing the blasticidin resistance gene from a constitutive promoter, therefore it was expected to contain cells resistant to the antibiotic. By day 26, samples that received both the RNP and the donor DNA (Samples 4-6) began to approach the number of live cells observed in the positive control. By day 29, cells in the sample that received neither the donor DNA nor the RNP (Sample 1) and cells that received either the donor DNA alone (Sample 2) or the RNP alone (Sample 3) were entirely eliminated by blasticidin selection. At the same time, each of the samples that received both the RNP and the donor DNA (Sample 4-6) reached comparable cell survival numbers to those of the positive control (Sample 7). On day 29, samples containing live cells (sample 4-7) were collected and transferred to 100 mm plates for panning selection for live cells and a portion of each sample was used for genomic DNA extractions.

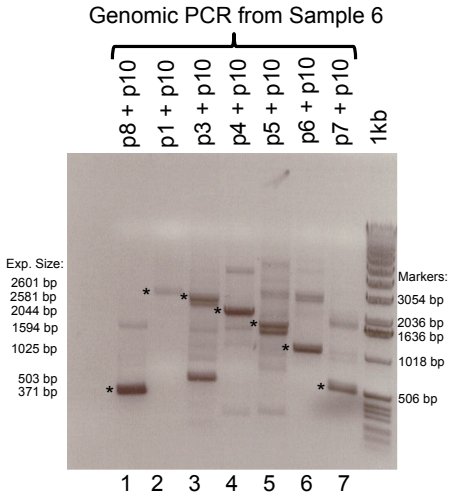
(a)


ATG (-) Bsd^R
pA
pUC19
-40 bp
M
pCCR5-PUF2
M
2.2 kb 5' HA
4.1 kb 3' HA
CS
A
A
501
1903
Exon 4
Integration
p1 p2 p3 p4 p5 p6 p7 p8 p9 p10
A
Bsd^R
501
1903
Exon 4
pA
pUC19
A
End of 5' HA
End of 3' HA

(b)


Sample #: 8 7 6 5 4
1 2 3 4 5

(c)


Genomic PCR from Sample 6
p8 + p10
p1 + p10
p3 + p10
p4 + p10
p5 + p10
p6 + p10
p7 + p10
1kb
Exp. Size:
2601 bp
2581 bp
2044 bp
1594 bp
1025 bp
503 bp
371 bp
Markers:
3054 bp
2036 bp
1636 bp
1018 bp
506 bp
1 2 3 4 5 6 7

(d)

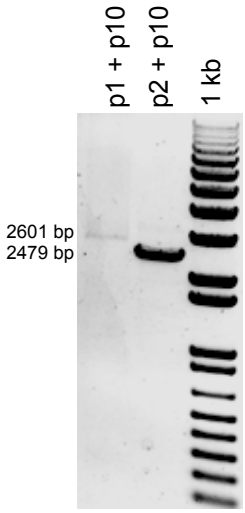

p1 + p10
p2 + p10
1 kb
2601 bp
2479 bp

Figure 3.9 Detection of CCR5 PUF Integration by PCR

Blasticidin-resistant cells that received the CCR5-targeted RNP and donor DNA were screened by PCR to detect PUF integrations. (a) A schematic diagram showing the PUF and the insertion product at the CCR5 target site. Primers designed to anneal inside the blasticidin resistance gene and at successively increasing distances from the insertion site are shown as arrows above the insertion product (p1-p10). Primers, p1 (CCR5 exon 1) and p2 (CCR5 exon 1b), were designed to anneal upstream of the 5'-homology region. Primers, p9 (Blasticidin 2 anti) and p10 (Blasticidin 1 anti), were designed to anneal within the blasticidin-resistance gene. PCR using primer sets, p1 and p10 and p2 and p9 allowed for nested PCR of the insertion event at the CCR5 target site. (b) Nested PCR products using primers p2 and p9 resulted in amplification of a non-specific band in sample 4 (Lane 5), sample 5 (Lane 4), sample 6 (Lane 3) and sample 8 (Lane 2) with a streak of non-specific amplification in the lane containing the PCR products from sample 6 (Lane 3). Closer inspection revealed a faint band corresponding to a 2.6-kb PCR product in the lane for sample 6 (Lane 3). This led to additional PCR analysis for sample 6. (c) Primers positioned within the CCR5 gene at successively increasing distances upstream of the integration site (p3, p4, p5, p6 and p7) were used in combination with the anti-parallel blasticidin resistance gene primer (p10, Blasticidin 1 anti) to amplify PCR products ranging from 503 bp to 2,601 bp. The expected sizes of the PCR products when using primer p10 (Blasticidin 1 anti) with the CCR5 primers are as follows: p7, CCR5 3639-3660 (503 bp product), p6, CCR5 3117-3136 (1,025 bp product), p5, CCR5 2548-2570 (1,594 bp product), p4, CCR5 2098-2121 (2,044 bp product), p3, CCR5 1561-1585 MluI (2581 bp product) and p1, CCR5 exon 1 (2,601 bp product corresponding to PUF insertion). Asterisks mark each of the amplification products on the gel (c). While each of the smaller PCR products could be amplified from randomly integrated donor DNA, the 2.6-kb fragment could only be amplified if the donor DNA integrated at the target site (Lane 2). (d) The primary PCR product (Lane 1) and the nested PCR product (Lane 2) are shown to be 2,601 and 2,479 bp, respectively. The nested PCR product was successfully sequenced to confirm the targeted integration.

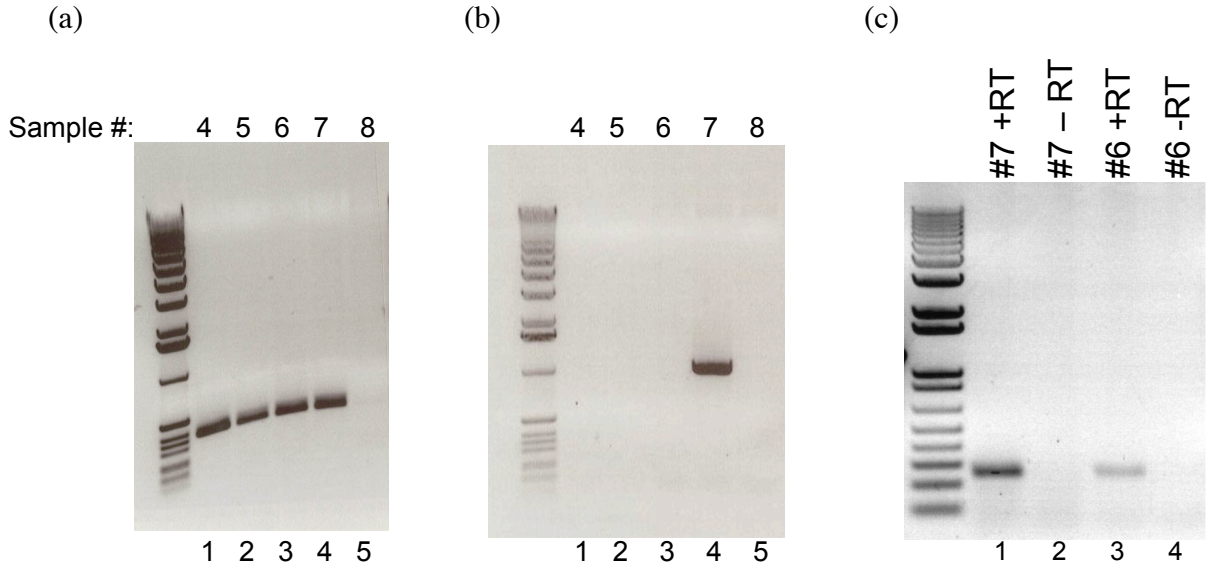
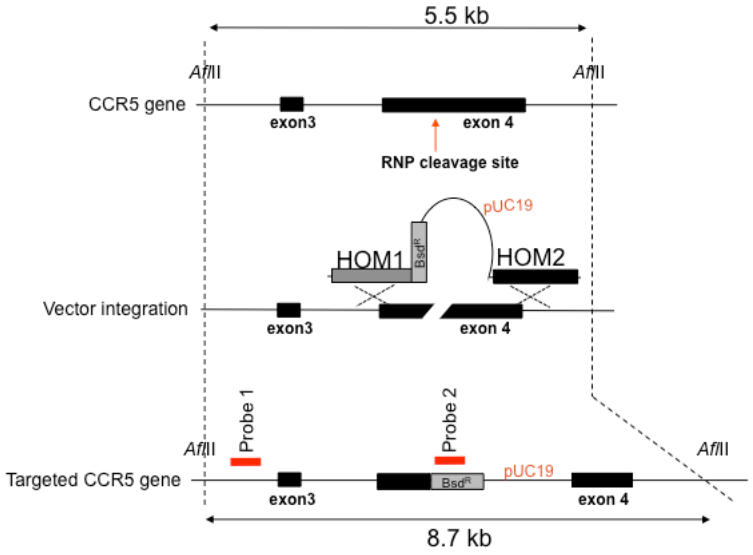


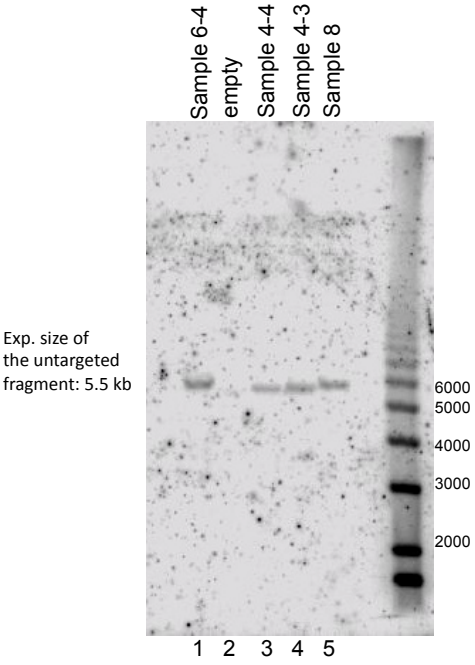
Figure 3.10 Additional PCR and RT-PCR to Characterize the Nature of the Blasticidin Resistance of K562 cells Treated with pCCR5-PUF2 and the CCR5-Targeted RNP

(a) To confirm that the blasticidin resistance gene was present in each of the samples containing blasticidin-resistant cells, PCR was performed using primers, XbaI Bsd –ATG and Blasticidin 1 anti, p8 and p10, according to Figure 3.9a, to amplify the blasticidin-resistance gene. The Bsd^R gene was amplified from all the samples that received the donor DNA but not sample 8 which is genomic DNA from untreated K562 cells. (b) To rule out contamination between the positive control sample 7 and samples 4, 5 and 6, PCR was performed to amplify the blasticidin-resistance gene along with the CMV promoter using primers, MluI CMV 5' sense and Blasticidin 1 anti. Since MluI CMV 5' sense anneals in the promoter sequence from pCMV/Bsd it should only amplify a product with the expression plasmid is present. Lane 4 shows the amplification product from pCMV/Bsd. Based on these results, none of the other samples contains the plasmid. (c) RNA was extracted from samples 6 and 7 for RT-PCR using the RNAqueous 4PCR Kit (Ambion), according to manufacturer's instructions. cDNAs were generated using the primer, Blasticidin 1 anti, and the Superscript II RT Kit (Invitrogen). The cDNAs generated from sample 6 and sample 7 with and without RT were used as templates for PCR. The Bsd^R ORF was amplified using primers, XbaI BSD –ATG and Blasticidin 2 anti. Lane 1 and Lane 3 contain the PCR product corresponding to expression of the blasticidin-resistance gene in samples 6 and 7.

(a)


5.5 kb
AflII
AflII
CCR5 gene
exon3
exon 4
RNP cleavage site
pUC19
HOM1
BsdR
HOM2
Vector integration
exon3
exon 4
Probe 1
Probe 2
AflII
AflII
Targeted CCR5 gene
pUC19
BsdR
exon3
exon 4
8.7 kb

(b)


Sample 6-4
empty
Sample 4-4
Sample 4-3
Sample 8
Exp. size of
the untargeted
fragment: 5.5 kb
6000
5000
4000
3000
2000
1 2 3 4 5

(c)

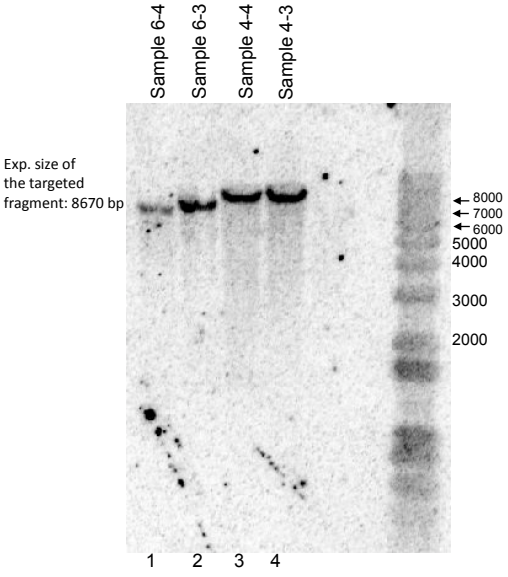

Sample 6-4
Sample 6-3
Sample 4-4
Sample 4-3
Exp. size of
the targeted
fragment: 8670 bp
8000
7000
6000
5000
4000
3000
2000
1 2 3 4

Figure 3.11 Southern Blot Strategy to Detect pCCR5-PUF2 Integration in Genomic DNA from CCR5-Targeted K562 Cells.

(a) Southern blot strategy to confirm the integration at the CCR5 target site. This strategy involved the use of two DNA probes. Probe 1 was designed to anneal to the CCR5 genomic sequence just upstream of the 5' end of the 5' homology arm. Probe 2 was designed to anneal to the blasticidin-resistance gene. Genomic DNA extracted from cells of sample 6 was to be digested with *Afl*II. The genomic DNA from non-targeted cells would result in a 5.5-kb DNA fragment, while successful integration would yield a DNA fragment 8.7 kb in length. The results of the Southern blot were complicated by the fact that the genomic DNA of sample 6 was extracted from a heterogeneous pool of cells collected prior to panning for isolated antibiotic-resistant cells. (b) When blotting against genomic DNA from sample 4, sample 6 and untreated K562 cells with the probe designed to anneal to the CCR5 gene just upstream of the targeting event, a band corresponding to the expected size of the integrated targeting vector was not detectable. The only band that was present in the Southern blot corresponded to the wild-type fragment. (c) Using the probe designed to anneal to the blasticidin resistance gene would verify the presence of the blasticidin ORF in the genomic DNA and offer another opportunity to confirm integration of the donor DNA. The Southern blot yielded a single band in each lane containing genomic DNA from blasticidin-resistant cells. Unfortunately, the single band was approximately 9 kb and was indistinguishable from the expected sized band for the linearized vector alone, which, albeit unlikely, may be present to some degree as extrachromosomal DNA (9.6 kb).

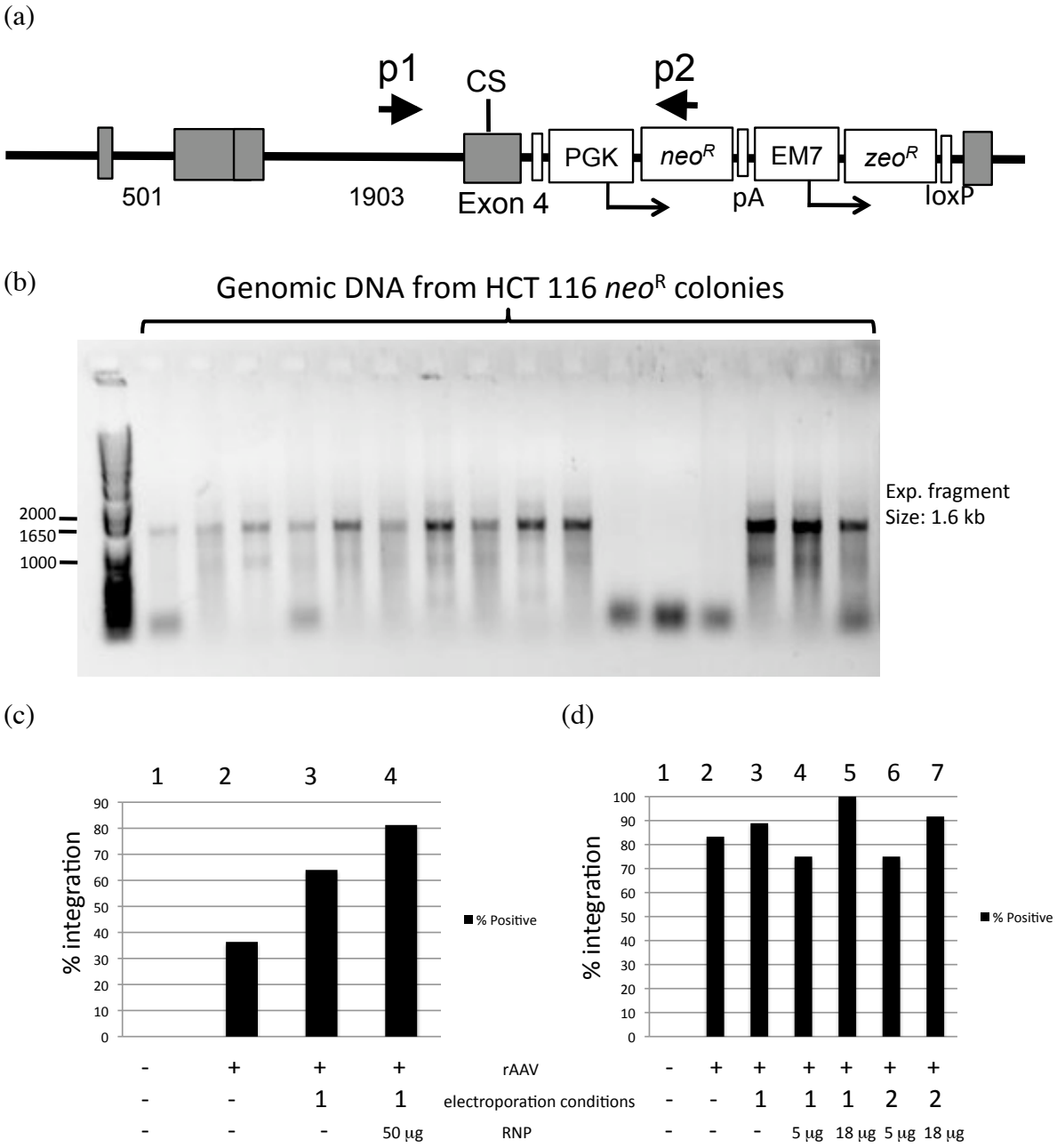


Figure 3.12 The CCR5-Targeted Group II Intron RNP Stimulates Integration of a rAAV Gene Targeting Vector at the CCR5 Locus

(a) PCR strategy to detect rAAV integration at the CCR5 target site. The PCR-based strategy for identifying cell clones that contain recombinant alleles involved the amplification of a DNA fragment using a primer that annealed in the CCR5 gene outside the homology region of the rAAV vector (p1) and a second primer that anneals within the Neo^R gene (p2). The resulting amplicon spans the entire HA and will only be amplified

when the marker is integrated at the target site. PCR were carried out in 20 μL reaction volumes using modified parameters from Rago et al. (2007). The PCR products were separated by gel electrophoresis and lanes were examined for the presence or absence of a PCR product of the correct size. (b) A typical PCR screening panel for rAAV targeting experiments. Because the homology arms used for targeting were ~900 bp, the complete amplicon, which consisted of the homology arm and a portion of the *neo*^{R} expression cassette totaled 1.6 kb. Presence of the correct sized PCR product indicated that the specific targeting event had occurred at the CCR5 target site. PCR products were sequenced for confirmation. (c) The results of targeting the CCR5 gene using pAAV-CCR5 and the CCR5-targeted RNP. Sample 1 received no virus and no RNP. Sample 2 received the virus but was not treated with RNP, nor was this sample electroporated. Sample 3 received virus and received an electroporation pulse without any RNP. Sample 4 was treated with virus and electroporated with 50 μg of the CCR5-targeted RNP. Samples 3 and 4, which received the electroporation pulse with or without the RNP, exhibited a higher targeting efficiency than sample 2, which did not receive the pulse. While electroporation alone increases the targeting efficiency 1.7-fold above that observed for the non-electroporated sample (from 36% to 64%), cells that received the CCR5-targeted RNP (50 μg) showed an even higher targeting efficiency (81.25%), 2.2-fold above that observed for the non-electroporated sample. The additional increase in targeting efficiency associated with the inclusion of the group II intron RNP in the targeting experiment suggests that the RNP is possibly having an enhancing effect on rAAV integration. We would expect that double-strand cleavage at the target site by the group II intron RNP would stimulate the integration of the rAAV. In this case, it might be argued that the group II intron RNP provides an increase of 120% over the targeting efficiency for cells that received the virus without electroporation or the RNP. Further, the targeting efficiency for sample 4 is elevated by 27% above that for sample 3 which received the virus and the pulse but no RNP. (d) Targeting the CCR5 gene using pAAV-CCR5t and the CCR5-targeted RNP. Two different amounts of RNP were delivered using two different electroporation conditions. Cells electroporated with the higher amounts of RNP (18 μg) showed increased targeting efficiencies (100% and 93.3%) over those that received no RNP or 5 μg of RNP (83.3% and 75%, respectively), regardless of the electroporation conditions used. The sample that was treated with the higher amount of RNP and electroporation condition 1 exhibited a higher targeting efficiency than the sample that received the same amount of RNP and electroporation condition 2. This suggests that electroporation condition 1 more efficiently delivered the RNP to the cells than electroporation condition 2. Both samples containing the higher amount of RNP, regardless of the electroporation condition, show a higher targeting efficiency than the sample that received no CCR5-targeted RNP. Similar to the previous experiment, the targeting results suggest that electroporation of the HCT116 cells with the RNP has a stimulatory effect on the integration of the rAAV, presumably due to double strand cleavage by the RNP at the target site.

CHAPTER 4: MATERIALS AND METHODS

4.1 MATERIALS AND METHODS FOR GROUP II INTRON-BASED GENE TARGETING IN EUKARYOTIC EMBRYOS

4.1.1 Recombinant Plasmids

pACD2 contains the 0.9-kb Ll.LtrB-ΔORF intron and flanking exons cloned downstream of a T7lac promoter in a pACYC184-based vector with a *cam*^{R} gene, and has the LtrA ORF cloned downstream of the 3′ exon (Guo et al. 2000, Karberg et al. 2001). The Ll.LtrB-ΔORF intron contains an additional phage T7 promoter inserted in intron domain IV (DIV) for use in plasmid-based DNA-integration assays.

pACD2, pACD-mitf(235s) were used for *in vitro* transcription of Ll.LtrB-ΔORF intron RNAs, which are then reconstituted with purified LtrA protein to form RNPs ("targetrons") for gene targeting (Guo et al. 2000). The latter plasmid contains a Ll.LtrB-ΔORF intron in which the EBS2, EBS1 and δ sequences were modified to be complementary to IBS2, IBS1, and δ′ sequences in the DNA target site of the zebrafish *mitf* gene. The donor plasmid used for the *mitf* targetron has a δ′ residue in the 3′ exon complementary to the retargeted δ residue in the intron RNA for optimal RNA splicing (Perutka et al. 2004).

pBRR3-ltrB, the target plasmid for intron-integration assays, contains the Ll.LtrB homing site (ligated exon 1 and 2 of the ltrB gene from position −178 upstream to +91 downstream of the intron-insertion site) cloned upstream of a promoterless *tet*^{R} gene in an Amp^R pBR322-based vector (Guo et al. 2000, Karberg et al. 2001). pBRR3-mitf and pBRR3-Golden ex1 are similar target plasmids in which the homing site has been altered

to contain the corresponding target sites for the mitf(235s) and Golden ex1 introns, respectively.

pIMP-1P, used for expression of the LtrA protein, contains the LtrA ORF cloned downstream of a tac promoter and ϕ10 Shine-Dalgarno sequence in the protein-expression vector pCYB2 (New England Biolabs, Ipswich, MA) (Saldanha et al. 1999). LtrA is expressed from this plasmid as a fusion protein with a C-terminal tag containing an intein-linked chitin-binding domain, enabling LtrA purification via a chitin-affinity column, followed by intein-cleavage.

Plasmid DNAs were prepared using the QIAgen line of plasmid Miniprep and Midiprep Kits, unless otherwise specified.

4.1.2 Transcription of Intron RNA and Self-Splicing Reaction

Ll.LtrB-ΔORF intron RNPs were reconstituted from *in vitro*-synthesized intron RNA and purified LtrA protein by modification of a method described previously (Saldanha et al. 1999). A precursor RNA containing the Ll.LtrB-ΔORF intron and flanking exons was transcribed with phage T7 RNA polymerase (Megascript T7 Kit; Ambion, Austin, TX) from the appropriate pACD-based donor plasmid (see above), which had been linearized with *Nhe*I. The resulting precursor RNAs containing the Ll.LtrB-ΔORF intron and flanking exon sequences were self-spliced in 1.25 M NH_4Cl, 50 mM $MgCl_2$, 50 mM Tris-HCl, pH 7.5 for 3 h at 37°C, then ethanol-precipitated and dissolved in distilled water.

To reconstitute Ll.LtrB-ΔORF RNPs, the self-spliced RNA (100 nM) was re-natured by heating to 55°C in 10 ml of pre-splicing buffer [450 mM NaCl, 5 mM $MgCl_2$,

40 mM Tris-HCl, pH 7.5] and slowly cooling to 30°C prior to addition of 200 nM of purified LtrA protein and further incubation for 30 min at 30°C (Saldanha et al. 1999). The resulting RNPs were collected by ultracentrifugation in a Beckman 50.2 Ti rotor at 145,000 × g for 16 h at 4°C and resuspended in 50 μl of HKM [10 mM KCl, 10 mM $MgCl_2$, and 40 mM HEPES, pH 8.0]. The RNP preparations typically contain 60–70% intron lariat RNA with the remainder being precursor RNA plus smaller amounts of linear intron and ligated exons, and they typically have a concentration of 1-2 mg RNA/ml based on OD_{260}.

4.1.3 Expression and Purification of LtrA

E. coli strains BL21(DE3) were transformed with the pImp-1P, and single colonies were selected on LB plates containing ampicillin (50 μg/mL). For starter cultures, a single colony was inoculated into 50 mL of LB medium containing ampicillin (50 μg/mL), and the cultures were shaken at 225 rpm at 37°C overnight. Two mL of the overnight culture was then inoculated into 1 L of SOB medium containing antibiotics (50 μg/mL ampicillin) in a 4-L Ehrlenmeyer flask and grown at 37°C in a rotary shaker (225 rpm) for 2–4 h, until OD_{595} was 0.4–0.8. For induction, IPTG was added to a final concentration of 0.5 mM, and the incubation was continued at 25°C for 5-6 h. After induction, cells were collected by centrifugation in a Beckman JA-14 rotor (2,455 × g for 10 min at 4°C), and washed with 50 mL of ice-cold S-150 [150 mM NaCl]. The washed cell pellet was resuspended in 50 mL of ice-cold S-500 [500 mM NaCl, 50 mM Tris-HCl (pH 8.0), 0.1 mM EDTA, 0.1% NP-40], and lysozyme (Sigma, St. Louis, MO) was added to a final concentration of 0.1 mg/ml. The cells were then lysed by three cycles of

freeze–thawing between –70°C and 25°C, followed by sonication (Branson 450 Sonifier, Branson Ultrasonics Inc., Danbury, CT; two 10 s bursts at amplitude 60, with at least 10 s between bursts). After sonication, insoluble material was removed by centrifugation in a Beckman JA-14 rotor (22,095 × g for 15 min).

For chromatography, the cleared lysate was diluted to 80 mL with CB [0.5 M NaCl, 50 mM Tris-HCl (pH 8.0), 0.1 mM EDTA] and loaded on a chitin column (New England Biolabs; 2.5 cm × 5 cm plastic column with a 10 mL bed volume), which had been washed with at least 100 mL of CB at a flow rate of 2.0 mL/min. After loading the lysate, the column was washed with at least 500 mL of CB, and then with 250 mL of S-750 [0.75 M NaCl, 50 mM Tris-HCl (pH 8.0), 0.1 mM EDTA]. After these washes, the LtrA–intein junction was cleaved by addition of 30 mL of 30 mM DTT in CB. The column flow was stopped, and the column was incubated with the DTT overnight at 4°C. The released LtrA protein was then recovered by washing the column with 20 mL of CB. The pooled fractions were dialyzed against CB containing 50% glycerol, thereby concentrating the protein 4–7-fold. Aliquots of the dialyzed protein were stored at -70°C.

4.1.4 *In vitro* Target-Primed Reverse Transcription Assays

TPRT reactions were carried out by incubating 1 μg of unlabeled target plasmid with 1 μg of reconstituted RNP particles in 20 μL of TPRT Buffer [10 mM KCl, 10 mM $MgCl_2$, and 50 mM Tris-HCl (pH 7.5)] containing dATP, dGTP, and dCTP (0.2 mM each), and 20 μCi of [α-^{32}P]dTTP (3000 Ci/mmol; New England Nuclear, Boston, MA). The reactions were initiated by addition of the RNP particles, and the mixtures were incubated for 30 min at 37°C. Products were analyzed in a 1% agarose gel, which was

dried and subsequently autoradiographed with a PhosphorImager (Molecular Dynamics, Sunnyvale, CA).

4.1.5 Zebrafish *In Vitro* Fertilization and Microinjection

All zebrafish were reared at 28.5°C under a 14 h/10 h light/dark cycle. Wild-type *D. rerio* strains were the AB line (Streisinger 1981). For plasmid targeting experiments, sperm and eggs were collected from adult AB males and females for *in vitro* fertilization, as described (Westerfield, 2000). For genome targeting experiments, males and females from the AB line were crossed with *nac*^{w2} fish to obtain heterozygous F_1 progeny (Lister et al. 1999). Fertilized eggs were incubated at room temperature for ~15 min prior to injection. For group II intron plasmid targeting assays, target plasmid (0.5 mg/ml in a solution containing 3.125 mM of each dNTP and specified amounts of $MgCl_2$, with 0.25% phenol red) and Ll.LtrB-ΔORF RNPs (0.1-1.0 mg/mL with 0.25% phenol red, 10 mM KCl, 10 mM $MgCl_2$, and 40 mM HEPES, pH 8.0) were injected separately into 10 to 25 one-cell embryos in Steinberg's medium (Steinberg 1957) using a pressure system (Picospritzer III; Parker Hannifin) with 20 psi output. The injection volumes were ~5-10 nL. A micromanipulator (MN-151, Narishige) was used to manipulate the injection needles. For genome targeting experiments, injections were performed as described above, except that the second injection was used to add Mg^{2+} without the target plasmid. Ll.LtrB-ΔORF RNPs modified to target the *mitf* gene (0.05-1.0 mg/mL with 0.25% phenol red in HKM buffer) were injected separately from specified amounts of $MgCl_2$ in 0.25% phenol red in HKM buffer. After injection of plasmid targeting reagents, the embryos were washed with Steinberg's medium, pooled in a single 1.5-ml Eppendorf

tube in 500 µl Steinberg's medium, and incubated for 1 h at 30°C, prior to plasmid DNA extraction and transformation of *E. coli*. Embryos receiving the genome targeting reagents were incubated at 25°C in Steinberg's solution for 48-72 h, which is the normal time frame for the appearance of melanin in melanophores. After 48 h, the embryos were screened for pigment mutations.

4.1.6 Nucleic Acid Extraction from Injected Zebrafish Embryos and Determination of Targeting Efficiency

To isolate nucleic acids following injection and incubation, the embryos were transferred to SNET lysis buffer [20 mM Tris-HCl, pH 8.0, 5 mM EDTA, 400 mM NaCl, 1% SDS (w/v), 400 µg/mL proteinase K (Molecular Biology Grade; Sigma-Aldrich)] for 1 h at 55°C, and then extracted twice with phenol-CIA. Nucleic acids were precipitated with isopropanol and dissolved in 20–50 µL of distilled water. For plasmid targeting experiments, two µL of the nucleic acid preparation was electroporated into *E. coli* HMS174(DE3) F^-, *hsdR*, *recA*, *rif*^{r} (Novagen, EMD Chemicals, Gibbstown, NJ), and cells were plated at different dilutions on Luria-Bertani (LB) medium containing ampicillin (50 µg/mL) plus tetracycline (25 µg/mL) or the same concentration of ampicillin alone. Colonies were counted after overnight incubation at 37°C, and the integration efficiency was calculated as the ratio of (Amp^R+Tet^R)/Amp^R colonies.

4.1.7 PCR Detection of Site-Specific Integration of Group II Introns in Plasmid and Genomic Targets in Zebrafish Embryos

For plasmid targeting, to detect intron integration by PCR for the WT RNP, the 3'-integration junction was amplified with the intron-specific primer, ltrB 890s, and a

plasmid-specific primer located downstream of the target site, Seq pBRR MCS+RR. The expected size of the amplicon was ~586 bp. For plasmid targeting using *mitf*-targeted RNPs, the 5'-integration junction was PCR amplified using the intron-specific primer, Seq pIMP EBS1/EBS2 (1), and a primer designed to anneal in the *mitf* cDNA of pBRR3-mitf upstream of the target site, mitfa/b upstream top. The expected size for the amplicon was 682 bp. For genomic targeting of the *mitf* gene, intron integration was screened by PCR amplification of the 5'-integration junction with the intron-specific primer, LtrB reverse bottom 1, and a *mitf*-specific primer located upstream of the target site, Upstream mitfa/b Top (1). The expected size of the amplicon was ~500 bp. For plasmid and genomic targeting, five μL of extracted DNA was used as the template for the PCR. The thermocycle program was as follows: 95°C for 5 min, followed by 30 cycles of 95°C for 30 sec, 55°C for 30 sec, 72°C for 2 min and final extension for 10 min at 72°C.

4.1.8 *Drosophila* Mating and Microinjection

The wild-type strain (Oregon-R) was grown in standard fly medium under standard conditions in the laboratory of Dr. David Stein. Embryos were collected on apple juice agar plates with added yeast paste for collection times not exceeding 60 min. All procedures following egg collection were performed at 18°C. The embryos were dechorionated with 50% bleach, rinsed in fresh water, adhered to a 24×30 mm glass cover slip, desiccated in a dish that contains Drierite and covered in 10S Voltalef Oil (VWR International). Microinjection was done using filament needles with a bore size of ~5 μm attached to an oil-filled 50-mL syringe. A micromanipulator was used to manipulate the injection needles. The injection volumes were ~0.3 nL.

The target plasmid DNA, pBRR3-ltrB, (0.3 mg/ml in dH_2O) was injected into the posterior of the embryo, followed within 5 min by the injection of WT RNPs (1.0 mg/ml), using different needles to avoid prior mixing. 70 embryos were injected for each condition and transferred to a humidified chamber for 30 min at 30°C. The embryos were collected with a pipette, pooled and transferred to a tube containing 100 μL of phenol-CIA for extraction. 50 μL of dH_2O was added to the extracted supernatant and the phenol-CIA extraction was repeated. The supernatant was mixed with 500 μL of SNET lysis buffer (see above) for 1 h at 55°C until the embryos were completely solubilized and then extracted once in phenol-CIA. Nucleic acids were precipitated with isopropanol and dissolved in 10 μl of distilled water. Plasmid DNAs were then transformed into *E. coli*, as described above for zebrafish assays.

4.1.9 Mobility Assay for the *mitf* Intron

For mobility assays, *E. coli* HMS174(DE3) (Novagen, Madison, WI) was cotransformed with pACD-mitf(235s) and pBRR3-mitf plasmids and grown in LB medium with chloramphenicol (25 μg/mL) and ampicillin (100 μg/mL) overnight (16 to 20 h, 37°C). A 50-μL sample of the overnight culture was inoculated into 10 mL of LB with the indicated antibiotics and grown to an OD_{595} of 0.2 to 0.3; at that point, 250 μL was inoculated into 5 mL of fresh LB without antibiotics and induced with 100 μM IPTG for 1 h at 37°C. Cells were harvested by centrifugation, washed, resuspended in 10 mL of ice-cold LB, and plated on LB plus ampicillin (50 μg/mL) in the presence or absence of tetracycline (25 μg/mL). Mobility frequencies were calculated as the proportion of $(Amp^R+Tet^R)/Amp^R$ colonies.

4.2 MATERIALS AND METHODS FOR GROUP II INTRON-BASED GENE TARGETING IN MAMMALIAN CELL CULTURE

4.2.1 Recombinant Plasmids

pACD-CCR5(332s) is a modified version of pACD2 in which the 0.9-kb Ll.LtrB-ΔORF intron has been modified such that EBS2, EBS1 and δ sequences are complementary to IBS2, IBS1, and δ′ sequences in the DNA target site for the human CCR5 gene (Guo et al. 2000). The Ll.LtrB-ΔORF intron contains an additional phage T7 promoter inserted in intron domain IV (DIV) for use in plasmid-based DNA-integration assays. The intron also has a δ′ residue in the 3′ exon complementary to the retargeted δ residue in the intron RNA for optimal RNA splicing (Perutka et al. 2004). pACD-CCR5(332s) is used for *in vitro* transcription of Ll.LtrB-ΔORF intron RNAs, which are then reconstituted with purified LtrA protein to form RNPs.

pBRR3-CCR5, the target plasmid for the CCR5(332s) intron, is a modified version of pBRR3-ltrB that contains the CCR5 homing site cloned upstream of the promoterless tet^R gene in an AmpR pBR322-based vector (Guo et al. 2000, Karberg et al. 2001).

pAAV-CCR5 and pAAV-CCR5t are modified versions of pAAV-MCS (Stratagene). pAAV-MCS is shuttle vector that contains AAV inverted terminal repeat (ITR) sequences necessary for AAV packaging plus *Not*I restriction sites useful for cloning the genetic cargo sequences. pAAV-CCR5 was constructed, as described (Rago, 2007, Kohli, 2004). pAAV-CCR5t was constructed specifically for use with the CCR5-targeted RNP. For this targeting vector, homology arms were PCR amplified from regions directly flanking the CCR5(332s) target site. The 5' homology arm corresponds

to nucleotides at positions -750 to -1 immediately upstream of the target site. The 3' homology arm corresponds to nucleotides +1 to +750 in relation to the intron insertion site. These fragments were cloned on either side of a selection marker cassette designed to confer resistance to neomycin (G418) and zeomycin. Dr. Ben Park at Johns Hopkins University prepared the selection cassette. Carlo Rago from the lab of Bert Vogelstein constructed pAAV-CCR5t, as described (Kohli et al. 2004).

pAAV-RC and pHelper are included with the AAV Helper-Free System (Stratagene). pAAV-RC contains the AAV-2 rep and cap genes expressed from two different promoters encoding the replication proteins and viral capsid structural proteins, respectively. pHelper contains the subset of adenovirus genes, VA, E2A and E4, which are necessary for high-titer AAV production in the AAV-293 cells.

pCMV/Bsd (Invitrogen) is a plasmid that expresses the blasticidin resistance gene using the bacterial EM7 promoter or the human CMV immediate-early promoter and contains a SV40 early polyadenylation signal sequence. This plasmid was used as a positive control for blasticidin selection in gene targeting experiments and as the template for PCR amplification of the ATG (-) Bsd^R open reading frame and SV40 polyA signal sequence for construction of the CCR5-targeted PUF.

4.2.2 TPRT Using CCR5-Targeted Group II Intron RNPs with Plasmid and Genomic DNA Targets

K562 cells were grown in Iscove's Modified Dulbecco's Medium (IMDM) with 10% FBS and penicillin/streptomycin. Cells were harvested by centrifugation for 5 min at 300 × g. Cells were resuspended in cold IMDM, counted and diluted to 5×10^5 cells in

10 mL. Cells were again centrifuged for 5 min at 300 × g. Genomic DNA was extracted using the GenElute Mammalian Genomic DNA Miniprep Kit (Sigma), according to the manufacturer's instructions. Genomic DNA concentrations were determined using a Nanodrop Microvolume Spectrophotometer (Thermo Scientific).

Targeting reactions were carried out by incubating 1 μg of the unlabeled plasmid DNA, pBRR3-CCR5, or specified amounts of genomic DNA with 1 μg of reconstituted CCR5-targeted RNP particles in 20 μL of TPRT Buffer [10 mM KCl, 10 mM $MgCl_2$, and 50 mM Tris-HCl (pH 7.5)] containing dATP, dGTP, dTTP and dCTP (0.2 mM each) (Invitrogen). The reactions were initiated by addition of the RNP particles, and the mixtures were incubated for 30 min at 37°C. After incubation, reactions were stopped by phenol-CIA extraction. The supernatant was transferred to a fresh clean tube and the DNA was ethanol precipitated. The DNA pellet was resuspended in 10-20 μL of TE [10 mM Tris-Cl, pH 7.5, 1 mM EDTA].

For *in vitro* plasmid targeting, to detect intron integration by PCR, the 3'-integration junction was amplified with the intron-specific primer, ltrB 816s, and a plasmid-specific primer located downstream of the target site, Seq pBRR MCS+RR. The expected size of the amplicon was 586 bp. To detect genomic DNA targeting, the 3'-integration junction was amplified by PCR with the intron-specific primer, ltrB 816s, and a CCR5-specific primer located downstream of the target site, CCR5 +367-+340. The expected size of the amplicon was ~492 bp. The thermocycle program designed for use with Taq DNA Polymerase (New England Biolabs) was as follows: 95°C for 5 min,

followed by 30 cycles of 95°C for 30 sec, 60°C for 30 sec, 72°C for 2 min and final extension for 10 min at 72°C.

4.2.3 Electroporation of K562 Cells with CCR5-Targeted RNPs and PCR Screening for Site-Specific Intron Integrations

K562 cells were grown in Iscove's Modified Dulbecco's Medium (IMDM) (Invitrogen) with 10% FBS and penicillin-streptomycin (Invitrogen) to 1×10^6 cells/mL. Cells were harvested by centrifugation at 300 × g for 5 min and washed 3 times with FBS (-) IMDM. The cell pellet was resuspended in an appropriate volume of FBS (-) IMDM to achieve a density of 6.25×10^6 cells/mL or 6.25×10^5 cells/mL. 5×10^6 cells (800 μL) or 5×10^5 cells (800 μL) were transferred to 1.5-mL microcentrifuge tubes and combined with specified amounts (10 μg or 20 μg) of CCR5-targeted RNPs at room temperature. 2 M $MgCl_2$ was added to some samples to a final concentration of 25 mM. Electroporations were immediately carried out at 320 V and 800 μF using a Gene Pulser II electroporation system (Bio-Rad Laboratories, Electroprotocols Bulletin #D035551). After electroporation, cells were incubated at either room temperature or 37°C for 15 min and transferred to a T-25 flask containing 10 mL of IMDM complete medium. Cells were incubated overnight at 37°C. Genomic DNA was extracted using the GenElute Mammalian Genomic DNA Miniprep Kit (Sigma).

Genomic DNA was screened for intron integrations by nested PCR amplification of the 3'-integration junction. For the primary PCR, I used the intron-specific primer, ltrB 816s, and a CCR5-specific primer, CCR5 +367-+340. Primary PCR products were purified using the QIAquick PCR Purification Kit (Qiagen). The purified PCR products

served as the template for nested PCR using primers, ltrB 890s and CCR5 +156-+140. The thermocycle program for the primary PCR was designed for use with Platinum Taq DNA Polymerase (Invitrogen) as follows: 94°C for 2 min, followed by 4 cycles of 94°C for 10 s, 64°C for 30 s, 68°C for 2 min, 4 cycles of 94°C for 10 s, 61°C for 30 s, 68°C for 2, 4 cycles of 94°C for 10 s, 58°C for 30 s, 68°C for 2 min and 30 cycles of 94°C for 10 s, 56.5°C for 30 s, 68°C for 2 min. The program used for nested PCR with Taq DNA Polymerase (New England Biolabs) was as follows: 95°C for 5 min, followed by 30 cycles of 95°C for 30 sec, 55°C for 30 sec, 72°C for 2 min and final extension for 10 min at 72°C.

4.2.4 CCR5 PUF Design and Construction

The CCR5 gene-targeting pick-up fragment (PUF) was constructed by cloning three separate PCR products into a pUC19 plasmid backbone. The three PCR fragments included the 5' homology arm corresponding to CCR5 genomic sequence beginning 30 nucleotides upstream of the CCR5(332s) intron target site, the 3' homology arm corresponding to CCR5 genomic sequence beginning 10 nucleotides downstream of the intron insertion site, and the ATG (-) blasticidin open reading frame with a SV40 polyA signal sequence. The 5' homology arm consisted of a 2.2-kb fragment amplified from genomic DNA of HEK293 cells, using primers CCR5 1561-1585 *Mlu*I and CCR5 3768-3744 *Xba*I. The 3' homology arm consisted of a 4.2-kb fragment also amplified from HEK293 genomic DNA, using primers CCR5 3803-3828 *Sal*I and CCR5 7991-7970 *Mlu*I. The blasticidin resistance gene was amplified without the start codon from pCMV/Bsd (Invitrogen), using primers *Xba*I Bsd –ATG and *Sac*I Bsd Dwnstrm polyA

signal. The PCR products were digested with the appropriate restriction enzymes and ligated into the pUC19 plasmid backbone pre-digested with *Sal*I and *Sac*I restriction enzymes. The restriction sites engineered into the PCR primers directed the orientation of the fragments in the PUF.

4.2.5 Electroporation and Selection of K562 Cells for RNP-Stimulated PUF Integration

K562 cells were grown in Iscove's Modified Dulbecco's Medium (IMDM), as previously described. $5x10^6$ cells were transfected by co-electroporation of 12.5 μg, 25 μg or 50 μg of CCR5-targeted RNP each with 10 μg of CCR5 PUF DNA linearized by *Mlu*I digestion. Control experiments included cells that were pulsed in medium alone and cells electroporated with 10 μg PUF alone, 25 μg RNP alone or 50 ng of pCMV/BSD (Invitrogen). The transfected cells were incubated for 15 min post-electroporation at room temperature and transferred to T-75 flasks containing 30 mL of IMDM without blasticidin. Cells were incubated at 37°C for 48 h. After the 48-h incubation, all the cell samples received IMDM containing blasticidin at a final concentration of 5 μg/mL. Cells were subsequently incubated at 37°C for 48 h. After 48 h with 5 μg/mL blasticidin, all cell samples received 30 mL IMDM containing blasticidin at a final concentration of 10 μg/mL. Subsequently, cells remained under constant selection in IMDM containing 10 μg/mL blasticidin for the next 25 days. From day 8 to day 29, cells were maintained in 10 mL of IMDM with 10 μg/mL blasticidin. On day 29, 5×10^6 cells each from samples 4-7 were collected by centrifugation at 300 × g for 5 min. The excess medium was removed and the cell pellets were stored at -70°C for genomic DNA extraction. Also on

day 29, 2×10^4 cells from each sample were diluted into 5 mL IMDM containing 10 μg/mL blasticidin and added to 100 mm culture plates that had been treated with WSC (Sigma) and ConA (Sigma), as described (Anderson & Junker 1994). Cells were grown on the surface of the plates for 4 days until colonies of cells were large enough to be picked up by 10 μL pipette tip. 5 days post-plating, colonies were picked from plates corresponding to each sample 4-7 and transferred to 96-well plates in 250 μL of IMDM with 10 μg/mL blasticidin. Cells were grown in 96-well plates for 4 days and transferred to 24 well plates containing IMDM with 10 μg/mL blasticidin. Cells were grown in 24 well plates for 4 days and transferred to T-25 flasks containing 5 mL of IMDM with 10 μg/mL blasticidin. Cells were grown in T-25 flasks until samples contained at least 2×10^7 cells for DNA and RNA extractions. Cells were harvested by centrifugation 300 × g for 5 min, medium was removed, and the cell pellets were stored at -70°C until DNA and RNA extractions were performed.

4.2.6 PCR Analysis of K562 Cells Treated with CCR5 RNPs and a Linearized dsDNA Plasmid PUF

Genomic DNA was extracted from frozen cell pellets of samples 4-7 and stock K562 cells (#8) using the GenElute Mammalian Genomic DNA Miniprep Kit (Sigma) for PCR analysis. Samples 4-7 were considered to contain heterogeneous populations of blasticidin-resistant cells because they were collected prior to the panning method of selection for isolated colonies, as described above. To confirm that the blasticidin resistance gene was present in each of the samples containing blasticidin-resistant cells, PCR was performed using primers, XbaI Bsd –ATG and Blasticidin 1 anti, to amplify the

blasticidin ORF sequence. To rule out contamination between the positive control sample 7 and the others, PCR was performed on samples 4-7 to amplify the blasticidin gene along with the CMV promoter using primers, MluI CMV 5' sense and Blasticidin 1 anti. To test the quality of the genomic DNA from sample 6 and to confirm the detectability of long PCR products, I designed a series of PCR that amplified fragments of increasing lengths when using the Blasticidin 1 anti primer with the following CCR5 primers: CCR5 3639-3660 (503 bp product), CCR5 3117-3136 (1,025 bp product), CCR5 2548-2570 (1,594 bp product), CCR5 2098-2121 (2,044 bp product), CCR5 1561-1585 MluI (2581 bp product), CCR5 exon 1 (2,601 bp product corresponding to PUF insertion). To confirm PUF integration, primers were designed upstream of the 5' homology region of the PUF and used with antisense blasticidin-resistance gene-specific primers for PCR to detect a 2.6-kb fragment (CCR5 exon 1 and Blasticidin 1 anti) and for nested PCR to detect a ~2.5-kb fragment (CCR5 exon 1b and Blasticidin 2 anti).

RNA was extracted from samples 6 and 7 for RT-PCR using the RNAqueous 4PCR Kit (Ambion), according to manufacturer's instructions. cDNAs were generated using the Blasticidin 1 anti primer and the Superscript II RT Kit (Invitrogen). The cDNAs from sample 6 and sample 7 with and without RT were used as templates for PCR. The Bsd^R ORF was amplified using primers, XbaI BSD –ATG and Blasticidin 2 anti. The thermocycle program for all PCR described here was as follows: 95°C for 5 min, followed by 30 cycles of 95°C for 30 sec, 55°C for 30 sec, 72°C for 2 min and final extension for 10 min at 72°C.

4.2.7 Southern Blot Analysis of K562 Genomic DNA to Detect PUF Integrations

Two DNA probes were PCR amplified for the Southern blotting experiments, a CCR5-specific probe and a Bsd^R gene-specific probe. The CCR5-probe was amplified from K562 genomic DNA using primers, CCR5 1267-1290 and CCR5 1510-1487. The Bsd^R gene probe was amplified from pCMV/Bsd (Invitrogen) plasmid DNA using primers, XbaI Bsd –ATG and Blasticidn 2 anti. The thermocycle program was designed for use with Phusion Mastermix as follows: 98°C for 10 sec, followed by 30 cycles of 98°C for 1 sec, 65°C for 5 sec, 72°C for 15 sec/1 kb and final extension for 1 min at 72°C. The PCR products were labeled using the High Prime DNA Labeling Kit (Roche), according to manufacturer's instructions. Two labeling reactions were performed for each probe. The two reactions were combined and unincorporated nucleotides were removed by eluting through Micro Bio-Spin P-30 spin chromatography columns (Bio-Rad).

For the Southern blot, genomic DNA was extracted from samples 4-7 and untreated K562 cells, as described above. Two to five μg of genomic DNA was digested overnight with restriction enzyme *Afl*II. The digested genomic DNA was phenol-CIA extracted twice and the digestion products were separated on a 1% agarose gel in 1X TAE at 100 V for 3 hours. The gel was stained with ethidium bromide and photographed on a UV transilluminator. The gel was subsequently soaked while slowly shaking in Denaturing Buffer for 30 min at room temperature, Neutralizing Buffer for 30 min at room temperature, and 2X SSC transfer buffer for 30 min at room temperature. The gel was rinsed with distilled water between each buffer exchange.

The DNA was transferred from the gel to a Nylon membrane (Ambion, Brightstar) using a homemade transfer apparatus. The transfer apparatus consisted of a glass dish with a glass plate covering most of the dish. 100 mL of 2X SSC was added to the dish. The plate was covered with a piece of filter paper folded so the ends touch the buffer (2X SSC). The gel was placed face down on the Whatman wick. A nitrocellulose membrane (Ambion, Brightstar) cut slightly larger than the gel was pre-soaked for 1 min in distilled water followed by 2 min in 2X SSC. The membrane was placed over the gel and the edges were covered with parafilm while not covering the wells or any lanes containing DNA. Two pieces of Whatman filter paper were pre-soaked in 2X SSC and placed over the parafilm and membrane. Four pieces of dry Whatman filter paper were placed on top and covered with full pack of dry paper towels followed by a 500 g weight for compression. The DNA transferred overnight at room temperature. The nitrocellulose membrane was removed and neutralized in 2X SSC for 5 min. While soaking in just enough 2X SSC to cover the membrane, the DNA was crosslinked to the membrane by placing in a Stratalinker on the "auto-crosslink" setting.

The nitrocellulose membrane with the freshly transferred genomic DNA was prepared for hybridization using Rapid-hyb buffer (Amersham), according to manufacturer's recommendations. The membrane was incubated in 15 mL Rapid-hyb Buffer (Amersham) in a closed hybridization tube for 30 min at 65°C while rotating. The probe was heated for 10 min in boiling water bath and added to 20 mL of fresh Hyb Buffer at 65°C. The Hyb Buffer in the hybridization tube was exchanged with the Hyb Buffer containing the probe and incubated for 2 h at 65°C. The membrane was washed

once for 20 min in 50 mL of 2X SSC, 0.1% SDS at 25°C, and twice in 50 mL 1X SSC, 0.1% SDS at 65°C. The membrane was then wrapped in saran wrap and autoradiographed.

4.2.8 Production of rAAV for Gene Targeting

rAAV were prepared two different ways. Either Carlo Rago prepared the virus, as previously described (Rago et al. 2007), or I prepared the virus according to a modified version of his protocol, as described here. Both protocols are modified versions of the AAV Helper-Free System Instruction Manual (Stratagene). AAV-293 cells (Stratagene) were grown to 70-80% confluency in 5 x 75-cm^2 flasks in DMEM (Invitrogen) supplemented with 10% FBS (Hyclone) and penicillin-streptomycin (Invitrogen). Cells were washed with HBSS (Invitrogen) and 7.5 mL of Opti-MEM (Invitrogen) was added. 50 μg of each plasmid pAAV-CCR5t, pAAV-RC (Stratagene) and pHelper (Stratagene) were combined in a 50-mL conical tube. Opti-MEM reduced-serum media (Invitrogen) was added to reach a total volume of 9.375 mL. In a separate tube, 375 μL of lipofectamine 2000 (Invitrogen) was diluted in Opti-MEM to a total volume of 9.375 mL. The DNA solution was added dropwise to the lipofectamine solution and incubated at room temperature for 15 min. 3.8 mL of the DNA-lipofectamine mixture was added dropwise to each of the flasks of cells and incubated at 37°C for 3-4 h. After the incubation, the Opti-MEM was replaced with DMEM (Invitrogen) supplemented with 10% FBS (Hyclone) and penicillin-streptomycin (Invitrogen), and the cells were incubated at 37°C for 48 h. The media was removed from the flask and the cells were scraped into 1 mL of PBS (Invitrogen) using a rubber

policeman. The cell suspension was transferred to a 15-mL tube and centrifuged at 1,000 × g for 5 min. The cells were resuspended in 3 mL of PBS and transferred to two microcentrifuge tubes. The cells were lysed by three cycles of freezing and thawing to release the virus. Each cycle consisted of 10 min in a dry ice-ethanol bath and 10 min at 37°C, vortexing after each thaw. The lysate was clarified by centrifugation at 15,000 × g in an Eppendorf desktop microcentrifuge 10 min at 4°C. Aliquots of the supernatant were stored at -70°C.

4.2.9 Electroporation and Selection of HCT116 Cells for Group II Intron RNP-Stimulated rAAV Integration

HCT116 cells were grown at 37°C to 60-80% confluency in HyClone McCoy's 5A medium in T-25 flasks. The growth medium was removed and the cells were washed with 5 mL of HBSS (Invitrogen). The HBSS was removed and specified volumes of rAAV were added dropwise to the cells. 2.5 mL of HyClone McCoy's 5A medium was added, and the cells were incubated for 2 h at 37°C. Each flask of cells was subsequently treated according to whether or not the cells would receive an electroporation pulse with or without the RNP. Cells receiving neither the pulse nor the RNP received 4 mL of growth medium and were incubated for 48 h at 37°C. Cells receiving the electroporation pulse were rinsed with 5 mL of HBSS. 1 mL of trypsin was added to the flask and incubated for 5-10 min at room temperature. 4 mL of complete medium was added to the trypsinized cells. The cells were harvested by centrifugation at 300 x g for 5 min. The cell pellet was resuspended in 5 mL of medium without FBS. The cells were washed twice in medium without FBS. The cells were counted and diluted in medium w/o FBS

to a final concentration concentration of 1.25×10^6 cells/mL. Specified amounts of RNPs were added to 800 μL aliquots of cells and the cells were electroporated under specified conditions. After electroporation, the cells were transferred to a T-25 flask containing 4 mL medium without FBS. After 2 h, the medium was exchanged for complete medium and incubated for 48 h at 37°C.

After 48 h, the cells were harvested from the T-25 flasks by trypsinization and resuspended in compete medium containing 400 μg/mL Geneticin (Invitrogen) to a final concentration of 1×10^5 cells/mL. 100 μL of the cell suspension for each sample was transferred to each well of a 96-well plate. The plates were sealed in plastic wrap and incubated at 37°C for 14-17 days. After the colonies were selected, 20 μL of trypsin was added to wells containing colonies. Once the cells were released from the surface of the plate, they were resuspended in 200 μL of complete medium and transferred into a second 96-well plate to consolidate all the colonies in a single plate. The new plates were incubated at 37°C until the cells reached 80% confluency. Genomic DNA was extracted from the 96-well plates using the 96-well Blood Kit (QIAgen), according to manufacturer's instructions. The purified DNA was stored at 4°C until needed for PCR analysis.

4.2.10 PCR Screening of HCT116 Cells for Site-Specific rAAV Integrations at the CCR5 Target Site

Genomic DNA was extracted colonies of HCT116 cells, as described above. The PCR strategy for screening the colonies was based on the amplification of the region upstream of the integration site using a primer designed to anneal 3' of the homology

region for the rAAV targeting vector, -897 CCR5, and a primer designed to anneal within the *neo*^{R} gene, NR Neo Reverse. 10 μL of the genomic DNA was used as the DNA template for the PCR. The thermocycle program for the primary PCR was designed for use with Platinum Taq DNA Polymerase (Invitrogen) as follows: 94°C for 2 min, followed by 4 cycles of 94°C for 10 s, 68°C for 30 s, 68°C for 2 min, 4 cycles of 94°C for 10 s, 64°C for 30 s, 68°C for 2, 4 cycles of 94°C for 10 s, 62°C for 30 s, 68°C for 2 min and 30 cycles of 94°C for 10 s, 58°C for 30 s, 68°C for 2 min. The PCR products were separated on a 1% agarose gel and targeting efficiencies were based on the presence or absence of a 1.6 kb amplicon.

Bibliography

Amsterdam, A., Burgess, S., Golling, G., Chen, W., Sun, Z., Townsend, K., et al. (1999). A large-scale insertional mutagenesis screen in zebrafish. *Genes & Development*, *13*(20), 2713-2724.

Andersen, A. S., & Junker, S. (1994). Simple and efficient recovery of rare living lymphoid cells from a vast majority of dead cells. *Nucleic Acids Research*, *22*(25), 5769-5770.

Bae, K., Kwon, Y. D., Shin, H., Hwang, M., Ryu, E., Park, K., et al. (2003). Human zinc fingers as building blocks in the construction of artificial transcription factors. *Nature Biotechnology*, *21*(3), 275-280.

Barkley, N. A., & Wang, M. L. (2008). Application of TILLING and EcoTILLING as Reverse Genetic Approaches to Elucidate the Function of Genes in Plants and Animals. *Current Genomics*, *9*(4), 212-226.

Beerli, R. R., & Barbas, C. F. (2002). Engineering polydactyl zinc-finger transcription factors. *Nature Biotechnology*, *20*(2), 135-141.

Belfort, M. (2003). Two for the price of one: a bifunctional intron-encoded DNA endonuclease-RNA maturase. *Genes & Development*, *17*(23), 2860-2863.

Blocker, F. J. H., Mohr, G., Conlan, L. H., Qi, L., Belfort, M., & Lambowitz, A. M. (2005). Domain structure and three-dimensional model of a group II intron-encoded reverse transcriptase. *RNA (New York, N.Y.)*, *11*(1), 14-28.

Bonen, L. (1993). Trans-splicing of pre-mRNA in plants, animals, and protists. *The FASEB Journal: Official Publication of the Federation of American Societies for Experimental Biology*, *7*(1), 40-46.

Bronson, S. K., & Smithies, O. (1994). Altering mice by homologous recombination using embryonic stem cells. *The Journal of Biological Chemistry*, *269*(44), 27155-27158.

Cathomen, T., & Joung, J. K. (2008). Zinc-finger nucleases: the next generation emerges. *Molecular Therapy: The Journal of the American Society of Gene Therapy*, *16*(7),

1200-1207.

Cavalier-Smith, T. (1991). Intron phylogeny: a new hypothesis. *Trends in Genetics: TIG*, *7*(5), 145-148.

Cech, T. R. (1986). RNA as an enzyme. *Scientific American*, *255*(5), 64-75.

Chandrasegaran, S., & Smith, J. (1999). Chimeric restriction enzymes: what is next? *Biological Chemistry*, *380*(7-8), 841-848.

Chen, Y., McClane, B. A., Fisher, D. J., Rood, J. I., & Gupta, P. (2005). Construction of an alpha toxin gene knockout mutant of *Clostridium perfringens* type A by use of a mobile group II intron. *Applied and Environmental Microbiology*, *71*(11), 7542-7547.

Cooley, L., Kelley, R., & Spradling, A. (1988). Insertional mutagenesis of the *Drosophila* genome with single P elements. *Science (New York, N.Y.)*, *239*(4844), 1121-1128.

Cornu, T. I., Thibodeau-Beganny, S., Guhl, E., Alwin, S., Eichtinger, M., Joung, J. K., et al. (2008). DNA-binding specificity is a major determinant of the activity and toxicity of zinc-finger nucleases. *Molecular Therapy: The Journal of the American Society of Gene Therapy*, *16*(2), 352-358.

Coros, C. J., Landthaler, M., Piazza, C. L., Beauregard, A., Esposito, D., Perutka, J., et al. (2005). Retrotransposition strategies of the *Lactococcus lactis* Ll.LtrB group II intron are dictated by host identity and cellular environment. *Molecular Microbiology*, *56*(2), 509-524.

Costa, M., Michel, F., & Westhof, E. (2000). A three-dimensional perspective on exon binding by a group II self-splicing intron. *The EMBO Journal*, *19*(18), 5007-5018.

Cousineau, B., Smith, D., Lawrence-Cavanagh, S., Mueller, J. E., Yang, J., Mills, D., et al. (1998). Retrohoming of a bacterial group II intron: mobility via complete reverse splicing, independent of homologous DNA recombination. *Cell*, *94*(4), 451-462.

Cui, X. (2005). *RNA/protein interactions during group II intron splicing and toward group II intron targeting in mammalian cells*. Ph.D. Dissertation, The University of Texas at Austin, Austin, TX.

Dai, L., & Zimmerly, S. (2003). ORF-less and reverse-transcriptase-encoding group II introns in archaebacteria, with a pattern of homing into related group II intron ORFs. *RNA (New York, N.Y.)*, *9*(1), 14-19.

Deiters, A., & Yoder, J. A. (2006). Conditional transgene and gene targeting methodologies in zebrafish. *Zebrafish*, *3*(4), 415-429.

Dellaporta, S. L., Xu, A., Sagasser, S., Jakob, W., Moreno, M. A., Buss, L. W., et al. (2006). Mitochondrial genome of *Trichoplax adhaerens* supports placozoa as the basal lower metazoan phylum. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(23), 8751-8756.

Deng, H., Liu, R., Ellmeier, W., Choe, S., Unutmaz, D., Burkhart, M., et al. (1996). Identification of a major co-receptor for primary isolates of HIV-1. *Nature*, *381*(6584), 661-666.

Deppenmeier, U., Johann, A., Hartsch, T., Merkl, R., Schmitz, R. A., Martinez-Arias, R., et al. (2002). The genome of *Methanosarcina mazei*: evidence for lateral gene transfer between bacteria and archaea. *Journal of Molecular Microbiology and Biotechnology*, *4*(4), 453-461.

Doetschman, T., Maeda, N., & Smithies, O. (1988). Targeted mutation of the *Hprt* gene in mouse embryonic stem cells. *Proceedings of the National Academy of Sciences of the United States of America*, *85*(22), 8583-8587.

Eickbush, T. H. (1999). Mobile introns: retrohoming by complete reverse splicing. *Current Biology: CB*, *9*(1), R11-14.

Eisen, J. S., & Smith, J. C. (2008). Controlling morpholino experiments: don't stop making antisense. *Development (Cambridge, England)*, *135*(10), 1735-1743.

Eskes, R., Liu, L., Ma, H., Chao, M. Y., Dickson, L., Lambowitz, A. M., et al. (2000). Multiple homing pathways used by yeast mitochondrial group II introns. *Molecular and Cellular Biology*, *20*(22), 8432-8446.

Eskes, R., Yang, J., Lambowitz, A. M., & Perlman, P. S. (1997). Mobility of yeast mitochondrial group II introns: engineering a new site specificity and retrohoming via full reverse splicing. *Cell*, *88*(6), 865-874.

Favard, C., Dean, D. S., & Rols, M. (2007). Electrotransfer as a non viral method of gene delivery. *Current Gene Therapy*, *7*(1), 67-77.

Fedorova, O., Mitros, T., & Pyle, A. M. (2003). Domains 2 and 3 interact to form critical elements of the group II intron active site. *Journal of Molecular Biology*, *330*(2), 197-209.

Frazier, C. L., San Filippo, J., Lambowitz, A. M., & Mills, D. A. (2003). Genetic manipulation of *Lactococcus lactis* by using targeted group II introns: generation

of stable insertions without selection. *Applied and Environmental Microbiology*, *69*(2), 1121-1128.

Galagan, J. E., Nusbaum, C., Roy, A., Endrizzi, M. G., Macdonald, P., FitzHugh, W., et al. (2002). The genome of *M. acetivorans* reveals extensive metabolic and physiological diversity. *Genome Research*, *12*(4), 532-542.

Gencheva, M., Anachkova, B., & Russev, G. (1996). Mapping the sites of initiation of DNA replication in rat and human rRNA genes. *The Journal of Biological Chemistry*, *271*(5), 2608-2614.

van Gent, D. C., Hoeijmakers, J. H., & Kanaar, R. (2001). Chromosomal stability and the DNA double-stranded break connection. *Nature Reviews. Genetics*, *2*(3), 196-206.

Golzio, M., Teissié, J., & Rols, M. (2002). Cell synchronization effect on mammalian cell permeabilization and gene delivery by electric field. *Biochimica Et Biophysica Acta*, *1563*(1-2), 23-28.

Gong, W. J., & Golic, K. G. (2003). Ends-out, or replacement, gene targeting in *Drosophila*. *Proceedings of the National Academy of Sciences of the United States of America*, *100*(5), 2556-2561.

Gregan, J., Kolisek, M., & Schweyen, R. J. (2001). Mitochondrial Mg^{2+} homeostasis is critical for group II intron splicing *in vivo*. *Genes & Development*, *15*(17), 2229-2237.

Guo, H., Karberg, M., Long, M., Jones, J. P., Sullenger, B., & Lambowitz, A. M. (2000). Group II introns designed to insert into therapeutically relevant DNA target sites in human cells. *Science (New York, N.Y.)*, *289*(5478), 452-457.

Guo, H., Zimmerly, S., Perlman, P. S., & Lambowitz, A. M. (1997). Group II intron endonucleases use both RNA and protein subunits for recognition of specific sequences in double-stranded DNA. *The EMBO Journal*, *16*(22), 6835-6848.

Hacein-Bey-Abina, S., Von Kalle, C., Schmidt, M., McCormack, M. P., Wulffraat, N., Leboulch, P., et al. (2003). LMO2-associated clonal T cell proliferation in two patients after gene therapy for SCID-X1. *Science (New York, N.Y.)*, *302*(5644), 415-419.

Hagedorn, M., Kleinhans, F. W., Artemov, D., & Pilatus, U. (1998). Characterization of a major permeability barrier in the zebrafish embryo. *Biology of Reproduction*, *59*(5), 1240-1250.

Hartwig, A. (2001). Role of magnesium in genomic stability. *Mutation Research*, *475*(1-2), 113-121.

Heap, J. T., Pennington, O. J., Cartman, S. T., Carter, G. P., & Minton, N. P. (2007). The ClosTron: a universal gene knock-out system for the genus *Clostridium*. *Journal of Microbiological Methods*, *70*(3), 452-464.

Hirata, R. K., & Russell, D. W. (2000). Design and packaging of adeno-associated virus gene targeting vectors. *Journal of Virology*, *74*(10), 4612-4620.

Hirata, R., Chamberlain, J., Dong, R., & Russell, D. W. (2002). Targeted transgene insertion into human chromosomes by adeno-associated virus vectors. *Nature Biotechnology*, *20*(7), 735-738.

Iversen, P. L., & Newbry, S. (2005). Manipulation of zebrafish embryogenesis by phosphorodiamidate morpholino oligomers indicates minimal non-specific teratogenesis. *Current Opinion in Molecular Therapeutics*, *7*(2), 104-108.

Jacquier, A., & Michel, F. (1987). Multiple exon-binding sites in class II self-splicing introns. *Cell*, *50*(1), 17-29.

Jones, J. P., Kierlin, M. N., Coon, R. G., Perutka, J., Lambowitz, A. M., & Sullenger, B. A. (2005). Retargeting mobile group II introns to repair mutant genes. *Molecular Therapy: The Journal of the American Society of Gene Therapy*, *11*(5), 687-694.

Karberg, M., Guo, H., Zhong, J., Coon, R., Perutka, J., & Lambowitz, A. M. (2001). Group II introns as controllable gene targeting vectors for genetic manipulation of bacteria. *Nature Biotechnology*, *19*(12), 1162-1167.

Kennell, J. C., Moran, J. V., Perlman, P. S., Butow, R. A., & Lambowitz, A. M. (1993). Reverse transcriptase activity associated with maturase-encoding group II introns in yeast mitochondria. *Cell*, *73*(1), 133-146.

Kimmel, C. B., & Law, R. D. (1985). Cell lineage of zebrafish blastomeres. I. Cleavage pattern and cytoplasmic bridges between cells. *Developmental Biology*, *108*(1), 78-85.

Kohli, M., Rago, C., Lengauer, C., Kinzler, K. W., & Vogelstein, B. (2004). Facile methods for generating human somatic cell gene knockouts using recombinant adeno-associated viruses. *Nucleic Acids Research*, *32*(1), e3.

Lambowitz, A. M., & Belfort, M. (1993). Introns as mobile genetic elements. *Annual Review of Biochemistry*, *62*, 587-622.

Lambowitz, A. M., & Zimmerly, S. (2004). Mobile group II introns. *Annual Review of Genetics*, *38*, 1-35.

Lechardeur, D., & Lukacs, G. L. (2002). Intracellular barriers to non-viral gene transfer. *Current Gene Therapy*, *2*(2), 183-194.

Ledwith, B. J., Manam, S., Troilo, P. J., Barnum, A. B., Pauley, C. J., Griffiths, T. G., et al. (2000). Plasmid DNA vaccines: assay for integration into host genomic DNA. *Developments in Biologicals*, *104*, 33-43.

Li, S., Xu, M., & Coelingh, K. (1995). Electroporation of influenza virus ribonucleoprotein complexes for rescue of the nucleoprotein and matrix genes. *Virus Research*, *37*(2), 153-161.

Lister, J. A., Robertson, C. P., Lepage, T., Johnson, S. L., & Raible, D. W. (1999). *nacre* encodes a zebrafish microphthalmia-related protein that regulates neural-crest-derived pigment cell fate. *Development (Cambridge, England)*, *126*(17), 3757-3767.

Liu, N., Caderas, G., Deillon, C., Hoffmann, S., Klauser, S., Cui, T., et al. (2001). Fusion proteins from artificial and natural structural modules. *Current Protein & Peptide Science*, *2*(2), 107-121.

Malik, H. S., & Eickbush, T. H. (1999). Modular evolution of the integrase domain in the Ty3/Gypsy class of LTR retrotransposons. *Journal of Virology*, *73*(6), 5186-5190.

Mandell, J. G., & Barbas, C. F. (2006). Zinc Finger Tools: custom DNA-binding domains for transcription factors and nucleases. *Nucleic Acids Research*, *34*(Web Server issue), W516-523.

Martin, W., & Koonin, E. V. (2006). A positive definition of prokaryotes. *Nature*, *442*(7105), 868.

Martinet, W., Schrijvers, D. M., & Kockx, M. M. (2003). Nucleofection as an efficient nonviral transfection method for human monocytic cells. *Biotechnology Letters*, *25*(13), 1025-1029.

Martínez-Abarca, F., García-Rodríguez, F. M., & Toro, N. (2000). Homing of a bacterial group II intron with an intron-encoded protein lacking a recognizable endonuclease domain. *Molecular Microbiology*, *35*(6), 1405-1412.

Matsuura, M., Noah, J. W., & Lambowitz, A. M. (2001). Mechanism of maturase-promoted group II intron splicing. *The EMBO Journal*, *20*(24), 7259-7270.

Matsuura, M., Saldanha, R., Ma, H., Wank, H., Yang, J., Mohr, G., et al. (1997). A bacterial group II intron encoding reverse transcriptase, maturase, and DNA endonuclease activities: biochemical demonstration of maturase activity and insertion of new genetic information within the intron. *Genes & Development*, *11*(21), 2910-2924.

Meister, G., Landthaler, M., Patkaniowska, A., Dorsett, Y., Teng, G., & Tuschl, T. (2004). Human Argonaute2 mediates RNA cleavage targeted by miRNAs and siRNAs. *Molecular Cell*, *15*(2), 185-197.

Michel, F., & Ferat, J. L. (1995). Structure and activities of group II introns. *Annual Review of Biochemistry*, *64*, 435-461.

Michel, F., & Jacquier, A. (1987). Long-range intron-exon and intron-intron pairings involved in self-splicing of class II catalytic introns. *Cold Spring Harbor Symposia on Quantitative Biology*, *52*, 201-212.

Michel, F., & Lang, B. F. (1985). Mitochondrial class II introns encode proteins related to the reverse transcriptases of retroviruses. *Nature*, *316*(6029), 641-643.

Michel, F., Umesono, K., & Ozeki, H. (1989). Comparative and functional anatomy of group II catalytic introns--a review. *Gene*, *82*(1), 5-30.

Miller, D. G., Petek, L. M., & Russell, D. W. (2003). Human gene targeting by adeno-associated virus vectors is enhanced by DNA double-strand breaks. *Molecular and Cellular Biology*, *23*(10), 3550-3557.

Mittal, V. (2004). Improving the efficiency of RNA interference in mammals. *Nature Reviews. Genetics*, *5*(5), 355-365.

Moehle, E. A., Moehle, E. A., Rock, J. M., Rock, J. M., Lee, Y., Lee, Y. L., et al. (2007). Targeted gene addition into a specified location in the human genome using designed zinc finger nucleases. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(9), 3055-3060.

Moens, C. B., Donn, T. M., Wolf-Saxon, E. R., & Ma, T. P. (2008). Reverse genetics in zebrafish by TILLING. *Briefings in Functional Genomics & Proteomics*, *7*(6), 454-459.

Mohn, A. & Koller, B.H. (1995) Genetic manipulation of embryonic stem cells. In: D.M. Glover and B.D. Hames, eds. *DNA Cloning 4-a Practical Approach: Mammalian Systems* (second ed.), IRL Press, Oxford.

Mohr, G., Smith, D., Belfort, M., & Lambowitz, A. M. (2000). Rules for DNA target-site

recognition by a lactococcal group II intron enable retargeting of the intron to specific DNA sequences. *Genes & Development*, *14*(5), 559-573.

Mondal, D., Williams, C. A., Ali, M., Eilers, M., & Agrawal, K. C. (2005). The HIV-1 Tat protein selectively enhances CXCR4 and inhibits CCR5 expression in megakaryocytic K562 cells. *Experimental Biology and Medicine (Maywood, N.J.)*, *230*(9), 631-644.

Mummidi, S., Ahuja, S. S., McDaniel, B. L., & Ahuja, S. K. (1997). The human CC chemokine receptor 5 (CCR5) gene. Multiple transcripts with 5'-end heterogeneity, dual promoter usage, and evidence for polymorphisms within the regulatory regions and noncoding exons. *The Journal of Biological Chemistry*, *272*(49), 30662-30671.

Müller, U. (1999). Ten years of gene targeting: targeted mouse mutants, from vector design to phenotype analysis. *Mechanisms of Development*, *82*(1-2), 3-21.

Nasevicius, A., & Ekker, S. C. (2000). Effective targeted gene 'knockdown' in zebrafish. *Nature Genetics*, *26*(2), 216-220.

Noah, J. W., & Lambowitz, A. M. (2003). Effects of maturase binding and Mg^{2+} concentration on group II intron RNA folding investigated by UV cross-linking. *Biochemistry*, *42*(43), 12466-12480.

Paddison, P. J., Caudy, A. A., & Hannon, G. J. (2002). Stable suppression of gene expression by RNAi in mammalian cells. *Proceedings of the National Academy of Sciences of the United States of America*, *99*(3), 1443-1448.

Palmer, J. D., & Logsdon, J. M. (1991). The recent origins of introns. *Current Opinion in Genetics & Development*, *1*(4), 470-477.

Patel, T., Bronk, S. F., & Gores, G. J. (1994). Increases of intracellular magnesium promote glycodeoxycholate-induced apoptosis in rat hepatocytes. *The Journal of Clinical Investigation*, *94*(6), 2183-2192.

Peebles, C. L., Perlman, P. S., Mecklenburg, K. L., Petrillo, M. L., Tabor, J. H., Jarrell, K. A., et al. (1986). A self-splicing RNA excises an intron lariat. *Cell*, *44*(2), 213-223.

Perutka, J., Wang, W., Goerlitz, D., & Lambowitz, A. M. (2004). Use of computer-designed group II introns to disrupt *Escherichia coli* DExH/D-box protein and DNA helicase genes. *Journal of Molecular Biology*, *336*(2), 421-439.

Plante, I., & Cousineau, B. (2006). Restriction for gene insertion within the *Lactococcus*

lactis Ll.LtrB group II intron. *RNA (New York, N.Y.)*, *12*(11), 1980-1992.

Podar, M., Chu, V. T., Pyle, A. M., & Perlman, P. S. (1998). Group II intron splicing *in vivo* by first-step hydrolysis. *Nature*, *391*(6670), 915-918.

Porteus, M. H., & Carroll, D. (2005). Gene targeting using zinc finger nucleases. *Nature Biotechnology*, *23*(8), 967-973.

Porteus, M. H., Cathomen, T., Weitzman, M. D., & Baltimore, D. (2003). Efficient gene targeting mediated by adeno-associated virus and DNA double-strand breaks. *Molecular and Cellular Biology*, *23*(10), 3558-3565.

Pruett-Miller, S. M., Connelly, J. P., Maeder, M. L., Joung, J. K., & Porteus, M. H. (2008). Comparison of zinc finger nucleases for use in gene targeting in mammalian cells. *Molecular Therapy: The Journal of the American Society of Gene Therapy*, *16*(4), 707-717.

Pyle, A. M., Fedorova, O., & Waldsich, C. (2007). Folding of group II introns: a model system for large, multidomain RNAs? *Trends in Biochemical Sciences*, *32*(3), 138-145.

Pyle, A. & Lambowitz, A.M. (2006) Group II Introns: ribozymes that splice RNA and invade DNA. In: Gesteland R.F., Cech T.R., Atkins, J.F. eds. *The RNA World*, 3rd ed. Cold Spring Harbor: Cold Spring Harbor Laboratory Press, 469-505.

Qin, P. Z., & Pyle, A. M. (1998). The architectural organization and mechanistic function of group II intron structural elements. *Current Opinion in Structural Biology*, *8*(3), 301-308.

Rago, C., Vogelstein, B., & Bunz, F. (2007). Genetic knockouts and knockins in human somatic cells. *Nature Protocols*, *2*(11), 2734-2746.

Ramirez, C. L., Foley, J. E., Wright, D. A., Müller-Lerch, F., Rahman, S. H., Cornu, T. I., et al. (2008). Unexpected failure rates for modular assembly of engineered zinc fingers. *Nature Methods*, *5*(5), 374-375.

Rest, J. S., & Mindell, D. P. (2003). Retroids in archaea: phylogeny and lateral origins. *Molecular Biology and Evolution*, *20*(7), 1134-1142.

Robertson, J. S. (1994). Safety considerations for nucleic acid vaccines. *Vaccine*, *12*(16), 1526-1528.

Rodriguez, S. A., Yu, J., Davis, G., Arulanandam, B. P., & Klose, K. E. (2008). Targeted inactivation of *Francisella tularensis* genes by group II introns. *Applied and*

Environmental Microbiology, *74*(9), 2619-2626.

Roitzsch, M., & Pyle, A. M. (2009). The linear form of a group II intron catalyzes efficient autocatalytic reverse splicing, establishing a potential for mobility. *RNA (New York, N.Y.)*, *15*(3), 473-482.

Romani, A. M., & Scarpa, A. (2000). Regulation of cellular magnesium. *Frontiers in Bioscience: A Journal and Virtual Library*, *5*, D720-734.

Romani, A. M. P., & Maguire, M. E. (2002a). Hormonal regulation of Mg^{2+} transport and homeostasis in eukaryotic cells. *Biometals: An International Journal on the Role of Metal Ions in Biology, Biochemistry, and Medicine*, *15*(3), 271-283.

Russell, D. W., & Hirata, R. K. (1998). Human gene targeting by viral vectors. *Nature Genetics*, *18*(4), 325-330.

Saldanha, R., Chen, B., Wank, H., Matsuura, M., Edwards, J., & Lambowitz, A. M. (1999). RNA and protein catalysis in group II intron splicing and mobility reactions using purified components. *Biochemistry*, *38*(28), 9069-9083.

Samulski, R. J., Chang, L. S., & Shenk, T. (1989). Helper-free stocks of recombinant adeno-associated viruses: normal integration does not require viral gene expression. *Journal of Virology*, *63*(9), 3822-3828.

San Filippo, J., & Lambowitz, A. M. (2002). Characterization of the C-terminal DNA-binding/DNA endonuclease region of a group II intron-encoded protein. *Journal of Molecular Biology*, *324*(5), 933-951.

Sandy, P., Ventura, A., & Jacks, T. (2005). Mammalian RNAi: a practical guide. *BioTechniques*, *39*(2), 215-224.

Schmelzer, C., & Schweyen, R. J. (1986). Self-splicing of group II introns *in vitro*: mapping of the branch point and mutational inhibition of lariat formation. *Cell*, *46*(4), 557-565.

Sedivy, J. M., & Dutriaux, A. (1999). Gene targeting and somatic cell genetics--a rebirth or a coming of age? *Trends in Genetics: TIG*, *15*(3), 88-90.

Segal, D. J., Beerli, R. R., Blancafort, P., Dreier, B., Effertz, K., Huber, A., et al. (2003). Evaluation of a modular strategy for the construction of novel polydactyl zinc finger DNA-binding proteins. *Biochemistry*, *42*(7), 2137-2148.

Shao, L., Hu, S., Yang, Y., Gu, Y., Chen, J., Yang, Y., et al. (2007). Targeted gene disruption by use of a group II intron (targetron) vector in *Clostridium*

acetobutylicum. *Cell Research*, *17*(11), 963-965.

Sharp, P. A. (1985). On the origin of RNA splicing and introns. *Cell*, *42*(2), 397-400.

Singh, N. N., & Lambowitz, A. M. (2001). Interaction of a group II intron ribonucleoprotein endonuclease with its DNA target site investigated by DNA footprinting and modification interference. *Journal of Molecular Biology*, *309*(2), 361-386.

Skelly, P. J., Hardy, C. M., & Clark-Walker, G. D. (1991). A mobile group II intron of a naturally occurring rearranged mitochondrial genome in *Kluyveromyces lactis*. *Current Genetics*, *20*(1-2), 115-120.

Smith, D., Zhong, J., Matsuura, M., Lambowitz, A. M., & Belfort, M. (2005). Recruitment of host functions suggests a repair pathway for late steps in group II intron retrohoming. *Genes & Development*, *19*(20), 2477-2487.

Steinberg, M.S. (1957) A non-nutrient culture medium for amphibian embryonic tissues. Carnegie Institute of Washington Year Book, 56, 347-348.

Streisinger, G, C Walker, N Dower, D Knauber, and F Singer. 1981. Production of clones of homozygous diploid zebra fish (*Brachydanio rerio*). *Nature* 291, no. 5813 (May 28): 293-296.

Toor, N., Hausner, G., & Zimmerly, S. (2001). Coevolution of group II intron RNA structures with their intron-encoded reverse transcriptases. *RNA (New York, N.Y.)*, *7*(8), 1142-1152.

Topaloglu, O., Hurley, P. J., Yildirim, O., Civin, C. I., & Bunz, F. (2005). Improved methods for the generation of human gene knockout and knockin cell lines. *Nucleic Acids Research*, *33*(18), e158.

Toro, N. (2003). Bacteria and Archaea Group II introns: additional mobile genetic elements in the environment. *Environmental Microbiology*, *5*(3), 143-151.

Toro, N., Jiménez-Zurdo, J. I., & García-Rodríguez, F. M. (2007). Bacterial group II introns: not just splicing. *FEMS Microbiology Reviews*, *31*(3), 342-358.

Toro, N., Molina-Sánchez, M. D., & Fernández-López, M. (2002). Identification and characterization of bacterial class E group II introns. *Gene*, *299*(1-2), 245-250.

Vallès, Y., Halanych, K. M., & Boore, J. L. (2008). Group II introns break new boundaries: presence in a bilaterian's genome. *PloS One*, *3*(1), e1488.

Vasileva, A., & Jessberger, R. (2005). Precise hit: adeno-associated virus in gene targeting. *Nature Reviews. Microbiology*, *3*(11), 837-847.

van der Veen, R., Arnberg, A. C., van der Horst, G., Bonen, L., Tabak, H. F., & Grivell, L. A. (1986). Excised group II introns in yeast mitochondria are lariats and can be formed by self-splicing in vitro. *Cell*, *44*(2), 225-234.

Vogel, J., & Börner, T. (2002). Lariat formation and a hydrolytic pathway in plant chloroplast group II intron splicing. *The EMBO Journal*, *21*(14), 3794-3803.

Wang, B., & Zhou, J. (2003). Specific genetic modifications of domestic animals by gene targeting and animal cloning. *Reproductive Biology and Endocrinology: RB&E*, *1*, 103.

Wank, H., SanFilippo, J., Singh, R. N., Matsuura, M., & Lambowitz, A. M. (1999). A reverse transcriptase/maturase promotes splicing by binding at its own coding segment in a group II intron RNA. *Molecular Cell*, *4*(2), 239-250.

Westerfield, M. (2000) *A guide for the laboratory use of zebrafish* (*Danio rerio*). Eugene: University of Oregon Press.

Yan, Z., Sun, X., & Engelhardt, J. F. (2009). Progress and prospects: techniques for site-directed mutagenesis in animal models. *Gene Therapy*, *16*(5), 581-588.

Yao, J., & Lambowitz, A. M. (2007). Gene targeting in gram-negative bacteria by use of a mobile group II intron ("targetron") expressed from a broad-host-range vector. *Applied and Environmental Microbiology*, *73*(8), 2735-2743.

Yao, J., Zhong, J., Fang, Y., Geisinger, E., Novick, R. P., & Lambowitz, A. M. (2006). Use of targetrons to disrupt essential and nonessential genes in *Staphylococcus aureus* reveals temperature sensitivity of Ll.LtrB group II intron splicing. *RNA (New York, N.Y.)*, *12*(7), 1271-1281.

Yáñez, R. J., & Porter, A. C. (1998). Therapeutic gene targeting. *Gene Therapy*, *5*(2), 149-159.

Yoon, Y., Sanchez, J. A., Brun, C., & Huberman, J. A. (1995). Mapping of replication initiation sites in human ribosomal DNA by nascent-strand abundance analysis. *Molecular and Cellular Biology*, *15*(5), 2482-2489.

Zhong, J., Karberg, M., & Lambowitz, A. M. (2003). Targeted and random bacterial gene disruption using a group II intron (targetron) vector containing a retrotransposition-activated selectable marker. *Nucleic Acids Research*, *31*(6), 1656-1664.

Zhong, J., & Lambowitz, A. M. (2003). Group II intron mobility using nascent strands at DNA replication forks to prime reverse transcription. *The EMBO Journal*, *22*(17), 4555-4565.

Zimmerly, S., Guo, H., Perlman, P. S., & Lambowitz, A. M. (1995). Group II intron mobility occurs by target DNA-primed reverse transcription. *Cell*, *82*(4), 545-554.

Zimmerly, S., Hausner, G., & Wu Xc. (2001). Phylogenetic relationships among group II intron ORFs. *Nucleic Acids Research*, *29*(5), 1238-1250.

Zimmerly, S., Moran, J. V., Perlman, P. S., & Lambowitz, A. M. (1999). Group II intron reverse transcriptase in yeast mitochondria. Stabilization and regulation of reverse transcriptase activity by the intron RNA. *Journal of Molecular Biology*, *289*(3), 473-490.

Vita

Jamie Lee Vernon was born in Concord, North Carolina to Elsie Mae Hathaway and George Brackett Vernon on December 12, 1973. He attended John T. Hoggard High School in Wilmington, North Carolina. In 1992, he entered North Carolina State University in Raleigh, North Carolina. During the summer of 1995 he attended the University of North Carolina at Wilmington. He received a Bachelor of Science from North Carolina State University in May 1997. Afterwards, he enrolled in classes at East Carolina University from 1997 through 1999. In August 1999, he began the Master of Science program in Molecular Biology/Biotechnology at East Carolina University. In 2001, he received the ECU, Department of Biology - Mary C. Helms Scholarship Award. He entered the Graduate School at the University of Texas at Austin in 2002. After working with Dr. Judy Edmiston as a teaching assistant, he received the 2003 UT Austin, School of Biological Sciences - Outstanding Teaching Award. While attending UT Austin, he completed the requirements for his Master's degree from ECU and remained a doctoral student in the Institute for Cell and Molecular Biology program. In 2004, he co-founded the ICMB, Paul D. Gottlieb Endowed Lecture Series, which has hosted Nobel Laureates James Watson, Sydney Brenner, Richard Roberts, Philip Sharp, Ferid Murad, Mario Capecchi and Jack Szostack on the UT Austin campus. During his 3rd year of graduate school, he was elected by his peers to serve as vice-president of the CMB graduate student association. In 2008, he co-founded the UT Austin Chapter of Scientists and Engineers for America, a non-profit organization that seeks to facilitate evidence-

based decision making at all levels of government. During the summer of 2009, he and his wife Cara were blessed with the birth of their daughter, Mia Nicole Vernon. He is co-inventor of one provisional patent application related to his dissertation research. He co-authored the following papers:

Mastroianni M., K. Watanabe, T.B. White, F. Zhuang, **J. Vernon**, M. Matsuura, J. Wallingford & A.M. Lambowitz. (2008) Group II intron-based gene targeting reactions in eukaryotes. PLoS ONE. 3(9): e3121.

Scemama, J.L., **J.L. Vernon**, E.J. Stellwag. Differential expression of *hoxa2a* and *hoxa2b* genes during striped bass embryonic development. (2006) Gene Expression Patterns 6(8): 843-8.

Vernon, J.L., P.C. Burr, J.E. Wiley and M. Farwell. Assignment of the mitochondrial translation elongation factor Ts gene (TSFM) to human chromosome 12 bands q13 --> q14 by *in situ* hybridization and with somatic cell hybrids. (2000) Cytogenetics and Cell Genetics (89): 145-146.

Permanent address: 7708 San Felipe Boulevard, Unit 41, Austin, TX 78729

This dissertation was typed by Jamie Vernon.